

---

# Creating Digital Editions for Corpus Linguistics

The case of *Potage Dyvers*,  
a family of six Middle English recipe collections

---

by

Ville Marttila

Department of Modern Languages  
University of Helsinki

*Academic dissertation to be publicly discussed,  
by due permission of the Faculty of Arts at the University of Helsinki  
in lecture room 6, on the 16th of August, 2014, at 10 o'clock.*

Helsinki 2014  
University of Helsinki

© Ville Marttila 2014

ISBN (paperback) 978-951-51-0059-7

ISBN (PDF) 978-951-51-0060-3

University of Helsinki

Unigrafia

Helsinki 2014

# Abstract

This thesis presents a corpus-linguistically oriented digital documentary edition of six 15<sup>th</sup>-century culinary recipe collections, known as the *Potage Dyvers* family, with an introduction to its historical context and an analysis of its dialectal and structural features, and defines an editorial framework for producing such editions for the purposes of corpus linguistic research. Traditionally historical corpora have been compiled from printed editions not originally designed to serve as corpus linguistic data. Recently, both the digitalisation of textual editing and the turning of corpus compilers towards original sources have blurred the boundaries between these two crafts, placing corpus compilers into an editorial role. Despite the fact that traditional editorial approaches have been recognised as largely incompatible with the needs of linguistic research, and the established methods of corpus encoding do not satisfactorily represent the documentary context of manuscript texts, no explicitly linguistic editorial approach has so far been designed for editing manuscript sources for use in corpora. Even most digital editions, despite their advanced representational capabilities, are literary or historical in orientation and thus do not provide an adequate model.

The editorial framework described here and the edition based on it have been explicitly designed to answer the needs of historical corpus linguistics. First, it aims at faithfully modelling the manuscript as a historical artefact, including both its textual content and its visual and material paratext, whose communicative importance has also been recognised by many historical linguists. Second, it presents this model in a form which allows not only the study of both text and paratext using corpus linguistic methods, but also allows resulting analytical metadata to be linked back to the edition, shared with other scholars, and used as the basis for further study. The edition itself is provided as a digital appendix to the thesis in the form of both a *digital data archive* encoded in TEI XML and three editorial presentations of this data, and serves not only as a demonstration of the editorial approach, but also provides a valuable new research resource.

The choice of material is based on the insight that utilitarian texts like recipes provide valuable material especially for historical pragmatics and discourse studies. As one of the first vernacular text types, recipes also provide an excellent opportunity to study the diachronic development of a single textual genre. The *Potage Dyvers* family is the second largest known family of Middle English recipe collections, surviving in six physically diverse manuscripts. Of these, four were edited in 1888 by conflating them into two collections, but their complex interrelationships have so far escaped systematic study. The structural analysis of the six *Potage Dyvers* versions indicates that the family, containing a total of 371 unique recipes, in fact consists of three sibling pairs of MSS. Two of these contain largely the same material but in a different order, while the third shares only a core of 89 recipes with the others, deriving a large number of recipes from other sources. In terms of their language, all of the six versions exhibit mainly Midlands forms and combine dialectally unmarked forms with more local variants from different areas, reflecting the 15<sup>th</sup>-century loss of dialectal distinctions which has not yet reached orthographic or morphological uniformity, and indicating possible metropolitan associations.



# Contents

<b>Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>xiii</b>
<b>List of Tables</b>	<b>xv</b>
<b>List of abbreviations</b>	<b>xvii</b>
<b>Preface</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The aim of this thesis . . . . .	5
1.2 Historical recipes as editorial material . . . . .	7
1.3 The structure of this thesis . . . . .	9
<b>I Theoretical background</b>	<b>13</b>
<b>2 Conceptualising the text</b>	<b>15</b>
2.1 Documents, texts, versions and works . . . . .	16
2.1.1 Document . . . . .	16
2.1.2 Text . . . . .	18
2.1.3 Version . . . . .	23
2.1.4 Work . . . . .	25
2.1.5 Material paratext or ‘bibliographic codes’ . . . . .	28
2.2 Context . . . . .	33
2.2.1 Layers of context . . . . .	33
2.2.2 Register and genre . . . . .	35
2.2.3 Context of historical texts . . . . .	37
2.2.4 The contextualized textual object . . . . .	38
2.3 Authors, scribes and textual transmission . . . . .	40
2.3.1 Conceptualising the author . . . . .	42
2.3.2 Production of manuscript texts . . . . .	43
2.3.3 Scribal copying . . . . .	44
2.4 Transcription of text . . . . .	48
2.5 Conclusion . . . . .	52

<b>3</b>	<b>Editing historical manuscripts</b>	<b>55</b>
3.1	Critical editing and textual criticism . . . . .	59
3.1.1	Theory and methodology of critical editing . . . . .	60
3.1.2	Critical editing of Middle English utilitarian manuscripts . . . . .	63
3.1.3	Implications of critical editing for linguistic research . . . . .	70
3.1.4	Critical editing and the digital medium . . . . .	75
3.2	Documentary editing . . . . .	77
3.2.1	Theory and methodology of documentary editing . . . . .	80
3.2.2	Criticism of documentary editing . . . . .	88
3.2.3	Documentary editing of multiple versions . . . . .	91
3.3	Conclusion . . . . .	96
<b>4</b>	<b>Digital editions for historical linguistics</b>	<b>99</b>
4.1	Philology and linguistics . . . . .	101
4.1.1	Theoretical background: New Philology . . . . .	101
4.1.2	Historical linguistics . . . . .	105
4.2	Scholarly digital editions . . . . .	111
4.2.1	Benefits of digital editions . . . . .	115
4.2.2	Criticism of digital editions . . . . .	125
4.3	Editing for historical corpus linguistics . . . . .	130
4.3.1	Compiling corpora from printed editions . . . . .	130
4.3.2	Manuscripts as historical linguistic witnesses . . . . .	136
4.3.3	Practical requirements of corpus linguistics . . . . .	141
4.3.4	Earlier linguistic editions . . . . .	151
4.4	Conclusion . . . . .	154
<b>5</b>	<b>Ontology of the digital edition</b>	<b>159</b>
5.1	Separation of data and its presentation . . . . .	161
5.1.1	Multiple identity of digital editions . . . . .	162
5.1.2	Benefits of separating data from its presentation . . . . .	164
5.1.3	Interoperability and standards . . . . .	165
5.2	Digital <i>archives</i> and digital <i>editions</i> . . . . .	166
5.3	From products to worksites . . . . .	170
5.3.1	Relationship between editor and user . . . . .	171
5.3.2	Digital editions as dynamic work-sites . . . . .	172
5.4	Modelling the document . . . . .	174
5.4.1	Text encoding and annotation as selective modelling . . . . .	176
5.4.2	Paratextual transcription . . . . .	176
5.4.3	Digitization as discrete sampling . . . . .	177
5.5	Economics of annotation . . . . .	180
5.5.1	Conflicting desiderata . . . . .	180
5.5.2	How much to annotate? . . . . .	182
5.5.3	Criteria for delimiting annotation . . . . .	183
5.6	Layered annotation . . . . .	186
5.6.1	Documentary annotation . . . . .	189
5.6.2	Descriptive annotation . . . . .	190
5.6.3	Analytical annotation . . . . .	191
5.6.4	Consequences of layered annotation . . . . .	193

5.7	The technology: XML and the TEI Guidelines . . . . .	198
5.7.1	eXtended Markup Language . . . . .	198
5.7.2	The Text Encoding Initiative and its Guidelines . . . . .	201
5.8	Conclusion . . . . .	209

## II Historical background 213

<b>6</b>	<b>Linguistic context:</b>	
	<b>multilingualism and literacy</b>	<b>215</b>
6.1	The trilingual environment of medieval England . . . . .	215
6.1.1	Changing relationships of England's languages . . . . .	218
6.1.2	The role of Latin in medieval England . . . . .	219
6.1.3	Role of French in medieval England . . . . .	220
6.2	Vernacularization and standardization . . . . .	223
6.2.1	Vernacularization of institutions . . . . .	224
6.2.2	Development of a standard . . . . .	226
6.3	Textual consequences of multilingualism . . . . .	231
6.3.1	Multilingual manuscripts . . . . .	231
6.3.2	Code-switching . . . . .	232
6.3.3	Linguistic borrowing . . . . .	235
6.4	Medieval literacies . . . . .	237
6.4.1	Defining medieval literacy . . . . .	240
6.4.2	Reading and writing . . . . .	241
6.4.3	Education and acquisition of literacy . . . . .	242
6.4.4	From aural to silent reading . . . . .	245
6.4.5	Changing functions of literacy . . . . .	247
6.4.6	Development of vernacular lay literacy . . . . .	249
6.5	Conclusion . . . . .	251
<b>7</b>	<b>Historical context of Middle-English recipes</b>	<b>253</b>
7.1	Cultural context: thought-styles regarding food . . . . .	254
7.1.1	Medical and dietary theories . . . . .	254
7.1.2	Religious doctrine . . . . .	264
7.1.3	Social factors . . . . .	268
7.2	Situational context: the medieval household . . . . .	273
7.2.1	The medieval kitchen and its staff . . . . .	273
7.2.2	From the great hall to the private chamber: the daily meal	278
7.2.3	A special occasion: the medieval feast . . . . .	280
7.3	Conclusion . . . . .	285
<b>8</b>	<b>Medieval culinary writing</b>	<b>289</b>
8.1	The European tradition of culinary writing . . . . .	290
8.1.1	Surviving English collections . . . . .	293
8.1.2	Existing editions and other research tools . . . . .	296
8.2	Medieval recipe collections as textual objects . . . . .	300
8.2.1	Recipe collections as <i>discourse colonies</i> . . . . .	302
8.2.2	Colonies within colonies: manuscript miscellanies . . . .	308

8.3	The recipe as a genre and a text type . . . . .	312
8.3.1	Textual characteristics of recipes . . . . .	313
8.3.2	Functional aspects of recipe as a <i>genre</i> . . . . .	320
8.3.3	Vagueness of medieval recipes . . . . .	324
8.3.4	Chronological evolution of medieval culinary recipes . . .	328
8.4	Producers and users of culinary recipes . . . . .	330
8.4.1	Recipe collections as evidence of historical practice . . . .	330
8.4.2	Questions of authorship . . . . .	333
8.4.3	Audience and function of medieval culinary recipes . . . .	337
8.5	Early printed recipe collections . . . . .	346
8.6	Conclusion . . . . .	347

### III The edition 349

9	The <i>Potage Dyvers</i> family of recipe collections	351
9.1	Identity of the <i>Potage Dyvers</i> family . . . . .	351
9.2	The manuscript versions . . . . .	353
9.2.1	London, British Library MS Harley 4016 . . . . .	355
9.2.2	London, British Library MS Harley 279 . . . . .	360
9.2.3	London, British Library MS Additional 5467 . . . . .	369
9.2.4	Oxford, Bodleian Library MS Ashmole 1439 . . . . .	379
9.2.5	Oxford, Bodleian Library MS Douce 55 . . . . .	385
9.2.6	Durham, Durham University Library MS Cosin V.iii.11 . .	391
9.3	Relationships to other collections . . . . .	403
9.3.1	English collections . . . . .	404
9.3.2	Continental collections . . . . .	411
10	Editorial principles	413
10.1	Conceptual structure of the edition . . . . .	416
10.1.1	Documentary annotation . . . . .	416
10.1.2	Descriptive annotation . . . . .	417
10.1.3	Analytical annotation . . . . .	420
10.1.4	Avoiding overlapping hierarchies . . . . .	423
10.2	Treatment of the text . . . . .	425
10.2.1	Delimiting the edited text . . . . .	426
10.2.2	Transcription of letterforms . . . . .	428
10.2.3	Capitalization and emphasis . . . . .	433
10.2.4	Abbreviation . . . . .	434
10.2.5	Punctuation . . . . .	439
10.2.6	Spacing . . . . .	440
10.2.7	Accuracy and proofreading . . . . .	440
10.2.8	Identification of scribal hands . . . . .	442
10.3	‘Non-editorial’ annotation . . . . .	443
10.3.1	Principles of linguistic annotation . . . . .	443
10.3.2	Intertextual annotation of parallel recipes . . . . .	450
10.4	Conclusion . . . . .	451



<b>11 Annotation practices</b>	<b>455</b>
11.1 Metadata header . . . . .	457
11.1.1 General annotation of metadata . . . . .	458
11.1.2 File description . . . . .	461
11.1.3 Encoding description . . . . .	464
11.1.4 Text profile . . . . .	465
11.1.5 Revision history . . . . .	467
11.1.6 Manuscript description . . . . .	467
11.2 Textual structure . . . . .	472
11.2.1 The front, back and body of a collection . . . . .	473
11.2.2 Recipe texts and textual divisions . . . . .	474
11.2.3 Headings, paragraphs, lists and items . . . . .	475
11.2.4 Arbitrary text segments . . . . .	477
11.2.5 Words, numbers, characters and punctuation . . . . .	477
11.2.6 Textual containers . . . . .	478
11.2.7 Linguistic identity of textual items . . . . .	480
11.3 Physical structure . . . . .	481
11.3.1 Foliation . . . . .	482
11.3.2 Columns . . . . .	482
11.3.3 Lineation . . . . .	483
11.3.4 Empty space . . . . .	484
11.3.5 Displaced items . . . . .	485
11.4 Genetic features of the text . . . . .	490
11.4.1 Scribal hands . . . . .	491
11.4.2 Emendation . . . . .	493
11.4.3 Manuscript notes and <i>forme work</i> . . . . .	498
11.5 Visual characteristics . . . . .	502
11.5.1 Highlighting, decoration and emphasis . . . . .	502
11.5.2 Images and ornaments . . . . .	504
11.5.3 Special symbols . . . . .	505
11.5.4 Differentiating special symbols and graphical elements . . . . .	511
11.6 Physical condition and legibility . . . . .	512
11.6.1 Damaged passages . . . . .	512
11.6.2 Unclear passages . . . . .	515
11.6.3 Missing text . . . . .	517
11.7 Annotation of abbreviation . . . . .	518
11.7.1 Abbreviated forms and abbreviation markers . . . . .	519
11.7.2 Editorial expansions . . . . .	520
11.7.3 Relating expansions to abbreviated words . . . . .	521
11.8 Editorial notes . . . . .	522
11.8.1 Textual notes . . . . .	523
11.8.2 Explanatory notes . . . . .	524
11.9 ‘Non-editorial’ analytical annotation . . . . .	526
11.9.1 Linguistic annotation . . . . .	527
11.9.2 Intertextual relationships . . . . .	531
11.9.3 Textual sources of the recipes . . . . .	533
11.9.4 Canonical titles . . . . .	533
11.10 Conclusion . . . . .	534

<b>IV Analysis</b>	<b>537</b>
<b>12 Dialectal characteristics</b>	<b>539</b>
12.1 Methods of analysis . . . . .	540
12.2 Dialectal features of the manuscript versions . . . . .	543
12.2.1 British Library MS Harley 4016 . . . . .	544
12.2.2 British Library MS Harley 279 . . . . .	548
12.2.3 British Library MS Additional 5467 . . . . .	553
12.2.4 Bodleian Library MS Ashmole 1439 . . . . .	556
12.2.5 Bodleian Library MS Douce 55 . . . . .	559
12.2.6 Durham University Library MS Cosin V.iii.11 A . . . . .	562
12.3 Shared features . . . . .	565
12.3.1 Dialectal similarity between versions . . . . .	565
12.3.2 Characteristics of shared forms . . . . .	569
12.4 Variety and degree of standardisation . . . . .	572
12.4.1 Distribution of variant forms . . . . .	573
12.4.2 Relationship to Chancery English and the emerging standard	584
12.5 Conclusion . . . . .	587
<b>13 Structural relationships</b>	<b>593</b>
13.1 Structural similarity of the versions . . . . .	595
13.2 Centrality and distribution of the recipes . . . . .	600
13.2.1 Comparing recipe order in pairs of versions . . . . .	606
13.3 Semantic organization of the versions . . . . .	609
13.3.1 MSS Ashmole 1439 and Harley 279 . . . . .	610
13.3.2 MSS Additional 5467 and Douce 55 . . . . .	611
13.3.3 MSS Cosin V.iii.11 A and Harley 4016 . . . . .	613
13.4 Conclusion . . . . .	615
<b>14 Conclusion</b>	<b>623</b>
14.1 Corpus-linguistic editions . . . . .	625
14.2 Cumulative production of knowledge . . . . .	627
14.3 The <i>Potage Dyvers</i> family . . . . .	631
<b>Bibliography</b>	<b>635</b>
<b>Appendices</b>	<b>681</b>
<b>A TEI XML Source Documents</b>	<b>681</b>
A.1 ODD and schema . . . . .	682
<b>B Diplomatic transcriptions</b>	<b>683</b>
B.1 Presentation conventions . . . . .	683
B.1.1 Scribal hands, emendations and annotations . . . . .	684
B.1.2 Highlighting and decoration . . . . .	684
B.1.3 Damage and legibility . . . . .	685
<b>C Parallel reading edition</b>	<b>687</b>

C.1	Presentation conventions . . . . .	687
C.1.1	Order of presentation . . . . .	688
C.1.2	Treatment of textual features . . . . .	688
C.1.3	Editorial notes . . . . .	689
C.2	Tables of recipes in each of the manuscripts . . . . .	690
C.3	Bills of fare included with the <i>Potage Dyvers</i> . . . . .	690
<b>D</b>	<b>Browsable HTML edition</b>	<b>693</b>
<b>E</b>	<b>XSL Transformations</b>	<b>695</b>
E.1	Generating the diplomatic transcriptions . . . . .	695
E.2	Generating the parallel presentation . . . . .	698
E.3	Generating the HTML edition . . . . .	699
<b>F</b>	<b>Detailed contents of the <i>Potage Dyvers</i> MSS</b>	<b>701</b>
<b>G</b>	<b>LALME analysis data</b>	<b>703</b>
G.1	Linguistic profiles . . . . .	703
G.2	Fully quantified linguistic profiles . . . . .	704
G.3	Exclusion templates for individual LALME forms . . . . .	705



# List of Figures

2.1	Variable centrality of discourse colony components . . . . .	27
2.2	Different levels of context and their influence on the textual object. . . . .	39
2.3	Different levels of manuscript copying. . . . .	48
3.1	Two models of textual transmission. . . . .	66
5.1	Relationships of different edition types. . . . .	168
5.2	An illustration of the relative economics of annotation. . . . .	181
5.3	Different layers of annotation. . . . .	188
7.1	Medieval food groups. . . . .	260
7.2	Humoral properties of some foodstuffs. . . . .	261
8.1	A sketch of a possible functional taxonomy of genres. . . . .	325
10.1	Conceptual structure of the edition. . . . .	417
11.1	The five principal regions denoted by the @place attribute. . . . .	487
11.2	The horizontal zones within a page region. . . . .	488
12.1	Examples of a dot map and an exclusion template. . . . .	543
12.2	Dialectal exclusion map for MS H4016. . . . .	545
12.3	Graduated dot map and a graduated exclusion map for MS H4016. . . . .	546
12.4	Dialectal exclusion map for MS H279. . . . .	548
12.5	Graduated dot map and a graduated exclusion map for MS H279. . . . .	550
12.6	A detail view of the dialectal maps for MS H279. . . . .	551
12.7	Dialectal exclusion map for MS Ad. . . . .	553
12.8	Graduated dot map and a graduated exclusion map for MS Ad. . . . .	555
12.9	Dialectal exclusion map for MS As. . . . .	557
12.10	Graduated dot map and a graduated exclusion map for MS As. . . . .	558
12.11	Dialectal exclusion map for MS D. . . . .	560
12.12	Graduated dot map and a graduated exclusion map for MS D. . . . .	561
12.13	Dialectal exclusion map for MS C. . . . .	563
12.14	Graduated dot map and a graduated exclusion map for MS C. . . . .	564
12.15	Dialectal exclusion map based on forms shared by all <i>PD</i> MSS. . . . .	571
12.16	Cumulative exclusion map of all <i>PD</i> MS versions. . . . .	571
12.17	Distribution of parallel forms for select LALME items in MS H4016. . . . .	575

12.18 Distribution of parallel forms for select LALME items in MS H279.	575
12.19 Distribution of parallel forms for select LALME items in MS Ad. .	578
12.20 Distribution of parallel forms for select LALME items in MS As. .	578
12.21 Distribution of parallel forms for select LALME items in MS C. . .	581
12.22 Distribution of parallel forms for select LALME items in MS D. . .	581
13.1 Structural similarity between different <i>PD</i> version pairs. . . . .	601
13.2 Transmissional relationships of the <i>PD</i> MSS suggested by quantitative analysis of the extent and ordering of recipes. . . . .	602
13.3 Edwards-Venn diagram of the recipes shared by different combinations of <i>PD</i> versions. . . . .	603
13.4 Transmissional relationships and postulated ancestors of the <i>PD</i> MSS suggested by the extent and ordering of recipes. . . . .	615
13.5 Subfamilies, works and versions within the <i>PD</i> family. . . . .	619
13.6 Distribution of recipes from different sources in the <i>PD</i> versions. .	621

# List of Tables

6.1	Relationships of the languages used in medieval England. . . . .	217
7.1	Medieval associations of the humors. . . . .	256
9.1	Overview of the physical features of the <i>PD</i> MSS. . . . .	354
9.2	Recipes shared by the <i>PD</i> MSS with the <i>OP</i> and <i>NBC</i> families. . .	407
12.1	<i>LALME</i> forms used for the dialectal analysis of MS H4016. . . . .	545
12.2	<i>LALME</i> forms used for the dialectal analysis of MS H279. . . . .	548
12.3	<i>LALME</i> forms used for the dialectal analysis of MS Ad. . . . .	553
12.4	<i>LALME</i> forms used for the dialectal analysis of MS As. . . . .	557
12.5	<i>LALME</i> forms used for the dialectal analysis of MS D. . . . .	560
12.6	<i>LALME</i> forms used for the dialectal analysis of MS C. . . . .	563
12.7	Examples of calculating similarity scores for <i>LALME</i> items. . . . .	567
12.8	Mean <i>LALME</i> item similarity scores for each <i>PD</i> MS pair. . . . .	568
12.9	<i>LALME</i> forms shared by all <i>PD</i> MS versions. . . . .	570
12.10	Summary of Chancery English features in the <i>PD</i> versions. . . . .	587
13.1	Quantitative structural similarities of the <i>PD</i> version pairs. . . . .	596
13.2	Number of shared and co-sequential recipes in different groups of three to six <i>PD</i> versions. . . . .	598
13.3	Recipes categorised according to their centrality to the <i>PD</i> family. .	602
13.4	Sequences of consecutive recipes shared by more than two MSS. . .	604
13.5	Differences in the extent of MSS C and H4016. . . . .	608
13.6	Sources of the <i>PD</i> family. . . . .	621





# List of abbreviations

## Manuscript sigla

Ad	London, British Library MS Additional 5467
As	Oxford, Bodleian Library MS Ashmole 1439
C	Durham, Durham University Library MS Cosin V.iii.11
D	Oxford, Bodleian Library MS Douce 55
H279	London, British Library MS Harley 279
H4016	London, British Library MS Harley 4016

## Acronyms

<i>ACOMESP</i>	<i>A Corpus of Middle English Scientific Prose</i>
ADE	Association for Documentary Editing
AND	<i>Anglo-Norman Dictionary</i>
ARCHER	<i>A Representative Corpus of Historical English Registers</i>
BL	British Library
BNC	<i>British National Corpus</i>
CD-ROM	Compact Disc Read-Only Memory
CEAA	Center for Editions of American Authors
CED	<i>Collins English Dictionary</i>
CES	Corpus Encoding Standard
CSS	Cascading Style Sheets
CTP	<i>Canterbury Tales Project</i>
DAF	<i>Dictionnaire de l'Académie française</i>
DALF	<i>Digital Archive of Letters in Flanders</i>
DC	<i>Diuersa Cibaria</i>

<b>DS</b>	<i>Diuersa Servisa</i>
<b>DTD</b>	Document Type Definition
<b>EAGLES</b>	Expert Advisory Group on Language Engineering Standards
<b>ECCO</b>	<i>Eighteenth-Century Collections Online</i>
<b>EEBO</b>	<i>Early English Books Online</i>
<b>EETS</b>	Early English Text Society
<b>EWD</b>	<i>English Witness Depositions 1560-1760: An Electronic Text Edition</i>
<b>FC</b>	<i>Forme of Cury</i>
<b>HC</b>	<i>Helsinki Corpus of English Texts</i>
<b>HIW</b>	<i>Henrik Ibsen's Writings</i>
<b>HTML</b>	HyperText Markup Language
<b>ISO</b>	International Organization for Standardization
<b>JITM</b>	Just-In-Time Markup
<b>KWIC</b>	KeyWord-In-Context
<b>LAEME</b>	<i>Linguistic Atlas of Early Middle English</i>
<b>LALME</b>	<i>Linguistic Atlas of Late Middle English</i>
<b>LAOS</b>	<i>Linguistic Atlas of Older Scots</i>
<b>LP</b>	linguistic profile
<b>ME</b>	Middle English
<b>MED</b>	<i>Middle English Dictionary</i>
<b>MEG-C</b>	<i>Middle English Grammar Corpus</i>
<b>MENOTA</b>	<i>Medieval Nordic Text Archive</i>
<b>MEP</b>	Model Editions Partnership
<b>MUFI</b>	Medieval Unicode Font Initiative
<b>NBC</b>	<i>A Noble Boke off Cookry</i>
<b>ODD</b>	One Document Does it All
<b>OED</b>	<i>Oxford English Dictionary</i>
<b>OHCO</b>	Ordered Hierarchy of Content Objects
<b>OLD</b>	<i>Oxford Latin Dictionary</i>
<b>OP</b>	<i>An Ordinance of Pottage</i>
<b>OSCE</b>	Open Source Critical Edition

<b><i>PD</i></b>	<i>Potage Dyvers</i>
<b>PDE</b>	Present-Day English
<b>PDF</b>	Portable Document Format
<b>POS</b>	Part-Of-Speech
<b>PUA</b>	Private Use Area
<b>SGML</b>	Standard Generalized Markup Language
<b>SVG</b>	Scalable Vector Graphics
<b>TEI</b>	Text Encoding Initiative
<b><i>UC</i></b>	<i>Utilis Coquinario</i>
<b>URI</b>	Uniform Resource Identifier
<b>URL</b>	Uniform Resource Locator
<b>WAB</b>	Wittgenstein Archives Bergen
<b><i>WWP</i></b>	<i>Women Writers Project</i>
<b>XCES</b>	XML Corpus Encoding Standard
<b>XHTML</b>	eXtensible HyperText Markup Language
<b>XML</b>	eXtensible Markup Language
<b>XSL</b>	eXtensible Stylesheet Language
<b>XSL-FO</b>	eXtensible Stylesheet Language Formatting Objects
<b>XSLT</b>	eXtensible Stylesheet Language Transformations



# Preface

Although doing research and writing a PhD is for the most part a solitary occupation, I have been fortunate enough to enjoy the company and support of a number of mentors and colleagues over the years I have spent preparing this thesis, for which it is my pleasure to offer my thanks. In terms of making this thesis possible, I am extremely grateful to the custodian libraries of the manuscripts edited in this thesis, namely The British Library, the Bodleian Library at the University of Oxford, and the Durham University Library, for granting me access to the manuscripts and to their staff for offering me expert help in locating any archival and catalogue sources regarding these manuscripts.<sup>1</sup>

I have also been fortunate to receive funding for my work from various institutions, including the Research Unit for Variation, Contacts, and Change in English (VARIENG), the Finnish Cultural Foundation, the Emil Aaltonen Foundation, and the University of Helsinki Faculty of Arts. My research has also been greatly aided by the travel grants provided by VARIENG, the Chancellor of the University of Helsinki, the Department of Modern Languages at the University of Helsinki, and the LANGNET graduate school, which have enabled not only the many library research trips necessary for the preparation of this thesis, but also my participation in numerous conferences that have kept me up to date with research in the various fields touched upon in this thesis, provided valuable feedback on my own research, and perhaps most importantly, brought me number of new colleagues and friends.

I would like to thank my esteemed external examiners, Prof. Emer. Graham Caie and Dr. Elena Pierazzo, for their insightful comments and valuable suggestions for improving my work in its final stages. Whatever shortcomings still remain are entirely my own. In addition to their duties as my examiners, they have both also served as personal influences to my career, Prof. Caie through his engaging teaching and encouraging attitude that pulled me further into the study of mediaeval English during my term as an undergraduate exchange student at the University of Glasgow in 2001, and Dr. Pierazzo through her inspirational work in the field of Digital Humanities as a long-time member and current chair of the TEI Council. I am also very grateful for the unfailing optimism and support of my supervisors, Prof. Irma Taavitsainen and Prof. Päivi Pahta, as well as for their acute comments and suggestions at various points during the research and writing process. The fact that I have managed to let go of my personal obsession with

---

<sup>1</sup> I am especially grateful to the eminent Dr. A. I. Doyle for providing me with invaluable unpublished information about the history of the DUL Cosin V.iii.11 manuscript.

completeness and actually finish this project owes much to their sage advice and persistence in making me heed it.

I would also like to acknowledge the fact that were it not for the uniquely inspiring environment provided by the VARIENG research unit, it is very unlikely that this work would have been undertaken in the first place. My gratitude for this goes first and foremost to Prof. Terttu Nevalainen, the Director of VARIENG during my stay there, and to Prof. Emer. Matti Rissanen, the original Director of the research unit and very much an embodiment of the spirit that made VARIENG such a special place, but also to all of the other people who served the research unit in various capacities, and to all the members of VARIENG who made it such a warm and hospitable environment over the years.

As a medievalist and manuscript scholar, I wish to give special thanks to Dr. Matti Kilpiö and Dr. Leena Kahlas-Tarkka, who not only share my interest in medieval manuscripts and have always been more than willing to share their considerable expertise with fledgling scholars like myself, but have also made all of our lives at VARIENG much brighter through their generosity and good spirit, regaling us with both song and wine on more occasions than I can count.

Within the extended family of VARIENG, my immediate academic family has been the Scientific Thought-styles project, founded and headed by my two supervisors. The members of this project have over the years become as much friends as colleagues. Dr. Martti Mäkinen was the senior postgraduate member of the project when I joined the team, just about to finish his PhD at the time. Not only a fellow medievalist, he is also my closest academic link to J.R.R. Tolkien, and the person who introduced me to the fantastic works of Robin Hobb, a somewhat dubious favour which I have subsequently passed on to several other people. Dr. Turo Hiltunen, frequently the voice of reason and moderation within the project team, has been an invaluable source of expertise and sound opinions on a variety of topics relating to corpus linguistic methodology. Apart from her academic prowess, Dr. Maura Ratia has always represented the spirit of fun in our project team, frequently being the *primus motor* behind the extracurricular activities accompanying our academic pursuits, and always excellent company. Dr. Carla Suhr, the resident historian of the project team and a fellow bibliophile with a penchant for historical fiction, has over the years become not only a beloved colleague but also a close friend and a favourite travel companion. Her sense of adventure and fun spirit are impressively matched by her many academic talents, including her uncanny organisational skills and the ability to prepare first-class conference presentations seemingly overnight. Dr. Jukka Tyrkkö, another fellow bibliophile, close friend and travel companion, seems to possess an unlimited supply of time and energy, which together with his almost intimidating intellect and many skills has resulted in a list of academic accomplishments that is as inspirational as it is humbling. The fact that his scholarly qualities are complemented by a genial nature and impressive social skills makes him excellent company on all social occasions, and goes a long way in explaining his vast network of friends and colleagues all over the world. Both Raisa Oinonen and Anu Lehto joined the project some years after me, following me as project assistants and quickly becoming an indispensable part of the team, both socially and academically.

The idea of creating digital editions for the purposes of corpus linguistics, from which a large part of this thesis springs, was originally formulated in late 2007 by

myself, Samuli Kaislaniemi and Dr. Alpo Honkapohja, all of whom were postgraduate students at VARIENG at the time. Although both of my colleagues have since redirected their attentions towards less digitally-oriented pursuits and the project founded to promote such editions—*Digital Editions for Corpus Linguistics* (DECL)—has since fallen more or less dormant, I would like to thank them for numerous fruitful—if at times contentious—discussions on the topic, which did much to shape my own thoughts on the issue.

In addition to all of the people already mentioned, I would also like to warmly thank the various current and former members of the VARIENG community with whom I have enjoyed many good times and stimulating conversations, including Simo Ahava, Dr. Alexandra Fodor, Bethany Fox, Mikko Hakala, Dr. Alaric Hall, Dr. Marianna Hintikka, Teo Juvonen, Henri Kauhanen, Minna Korhonen, Salla Lähdesmäki, Joe McVeigh, Dr. Anneli Meurman-Solin, Dr. Minna Nevala, Sara Norja, Dr. Arja Nurmi, Ulla Paatola, Prof. Minna Palander-Collin, Dr. Anni Sairio, Dr. Maija Stenvall, Tanja Säily, Tuuli Tahko, Dr. Heli Tissari, Turo Vartiainen and Anna-Liisa Vasko from Helsinki, and Dr. Alicia Jinkerson, Henna Jousmäki, Dr. Samu Kytölä, Dr. Leila Kääntä, Dr. Mikko Laitinen, Prof. Sirpa Leppänen, Terhi Paakkinen, Saija Peuronen, Prof. Arja Piirainen-Marsh, Prof. Anne Pitkänen-Huhta, Dr. Tiina Räisänen, Marianne Toriseva, and Dr. Elina Westinen from Jyväskylä.

My two-year stint as a self-funded member of the national Langnet postgraduate school also brought me a number of wonderful colleagues from outside the English-language bubble, namely Riikka Ala-Risku, Maija Hirvonen, Hanna Lantto, Heini Lehtonen, Dr. Kaarina Mononen, Meri Päivärinne and Max Wahlström, whom I would like to thank for providing not only fresh insights into the differences and commonalities in the linguistics of different languages, but also for their excellent company and many fond memories of shared seminars.

In addition to the Department of Modern Languages at the University of Helsinki, I have also had the pleasure of sharing scholarly insights with various wonderful people in other departments of the University of Helsinki and other Finnish universities, of whom I would specifically like to mention Dr. Jesse Keskiaho and Dr. Sini Kangas, fellow mediaevalists from the Department of Philosophy, History, Culture and Art Studies at the University of Helsinki, Rev. Dr. Ruth Carroll—one of the few historical linguists who have shared my interest in the language of culinary recipes—Prof. Risto Hiltunen, Dr. Matti Peikola, Hanna Salmi, Dr. Janne Skaffari, Dr. Sanna-Kaisa Tanskanen, and Mari-Liisa Varila from the English Department at the University of Turku, Prof. Anthony Johnson from Åbo Akademi, and Prof. Tapio Seppänen, Ilkka Juuso and Tuomo Toljamo from the University of Oulu. The last year of work on this thesis was shadowed by the untimely passing in 2013 of Prof. Lisa-Lena Opas-Hänninen from the University of Oulu, in whom I—along with many others—lost an extraordinary Digital Humanities scholar, a wise mentor, and a valued friend. I hope this thesis will to some small degree help carry on her legacy.

The steady flow of international visitors at VARIENG and the chance to regularly attend international conferences have also given me wonderful opportunities for meeting and enjoying stimulating conversations—both academic and social—with a large number of international colleagues. Of all these wonderful people, special mention has to be made of Dr. James Cummings, who has not only be-

come a personal friend, but whose expert advice in the initial stages of preparing the present edition had a significant influence on the development of its annotation practices. Similarly, I would also wish to express my special thanks to those members of the Glasgow University English Department whose classes I had the privilege of attending during my stay there as an undergraduate exchange student in 2001, for cultivating my interest in the history of English and all things medieval which ultimately lead to this thesis: Prof. Emer. Graham Caie, Prof. Simon Horobin, Prof. Carole Hough, Dr. Kathryn Lowe, and Prof. Jeremy Smith. Although I admit, in the words of the eminent adventurer Bilbo Baggins, that “I don’t know half of you half as well as I should like”, I would nevertheless like to thank the following colleagues from around the world for allowing me to enjoy their pleasant company and benefit from their expertise on various occasions: Ms. Juulia Ahvensalmi, Dr. Marc Alexander, Ms. Jean Anderson, Prof. Dawn Archer, Dr. Alistair Baron, Mr. David Beavan, Prof. Douglas Biber, Ms. Melanie Borchers, Ms. Mila Chao, Ms. Madalina Chitez, Mr. Øyvind Eide, Dr. Fátima Faya, Ms. Heather Froehlich, Ms. Kirsten Gather, Dr. Johanna Green, Prof. Sebastian Hoffmann, Prof. Andreas Jucker, Dr. Ágnes Kiricsi, Dr. Joanna Kopaczkyk, Prof. William Kretzschmar, Ms. Mareike Laue, Dr. Elaine Leong, Dr. Ursula Lutzky, Ms. Elizabeth MacDonald, Ms. Francesca Mackay, Dr. Teresa Marques Aguado, Dr. Stephen Morrison, Mr. Mike Olson, Dr. Goran Proot, Ms. Paula R. Puente, Dr. Paul Rayson, Prof. Antoinette Renouf, Prof. Jacob Thaisen, Mr. Edward Vanhoutte, and Dr. Nuria Yáñez-Bouza.

In addition to my academic colleagues, I would also like to give my heartfelt thanks to those friends who reside outside of my particular academic bubble—many of whom are also distinguished scholars in their own respective fields—for their continued support and understanding over my chronic lack of free time and energy over the years of writing this thesis, as well as for the provision of much-needed fun and games along the way. My oldest and closest friends, Kaapo, Puukko, Tobeliuz, Vainio, Mursu and Sami have persistently continued to invite me along to fun things even when I’ve been able to join them only occasionally. Hopefully I’ll make it to Lapland next year! My informally adopted big sister Ranja deserves the warmest of thanks for helping to keep my head screwed on even during the darker moments of writing this thesis, and for her help in keeping my social life alive. I would also like to thank the members of Poikkitieteellinen klubi, our quasi-academic secret society, who have provided a refreshing venue for deeply analytical discussions on a wide variety of topics in an enjoyable and informal setting. All of my wonderful friends and kindred spirits from the Finnish LARP, RPG and historical re-enactment scenes—some of whom have been around for twenty years, and some of whom have become dear friends in a matter of weeks—deserve my gratefulness for sharing my passion for all things fantastic, balancing my life with suitable doses of escapism, and for being the greatest of people to spend time with.

My family, despite their occasional good-natured inquiries as to whether I will some day be done with my studies, have always been very supportive of my academic pursuits. While my younger brother Olli has always had quite different interests and does not share my bookish nature, his career as a chef has nevertheless led us to a mutual interest in and respect for culinary culture, although from very different angles. I would like to thank my parents, Raija and Risto, for always



encouraging my inquisitive mind and allowing me freedom to be the bookish child I was, and for providing me with both emotional and material support on the occasions that I have needed it. Finally, I would like to dedicate this work to my beloved, long-suffering wife Meira, whose support and patience have exceeded all measure, and without whom I would not have made it through this ordeal. Thank you for your love and for keeping me sane.

Ville Marttila  
Helsinki 21.7.2014



# Chapter 1

## Introduction

Linguistic *corpora*—commonly defined as large and structured sets of texts—have traditionally focused on language as disembodied sequences of written or spoken utterances, represented in digital form as linear streams of character data. This is only natural, considering that the first English-language corpora were intended for the syntactic and lexical analysis of Present Day English texts, and the material aspects of texts were not relevant for these aims. Furthermore, the first corpora—compiled in the 1960s—were severely restricted by the technological limitations of the time,<sup>1</sup> and the representation of textual, typographical or bibliographical features beyond the indication of basic textual boundaries was simply not possible, especially since any metadata capabilities were reserved for the primary purpose of linguistic annotation. The conventions used in these early corpora and their exclusive focus on ‘pure text’ became the norm and also influenced the design of subsequent corpora. When the first English historical corpus, the *Helsinki Corpus of English Texts* (HC, see HC Manual), was compiled in the 1980s, it bore a strong family resemblance to its predecessors, including its focus on the text as a linear stream of character data.<sup>2</sup>

More recently this persistent emphasis on ‘the text itself’ has been challenged by new trends in historical linguistics, emphasising the importance of studying historical language use in its original *context*. Since the situational context and much of the cultural context of historical texts is inaccessible to us, the importance of the *documentary* context is relatively much greater than for present-day texts, both because it is all we have, and because the pragmatic functions of its material features are less well known. While the importance of various *paratextual* features, including the material paratext constituted by the physical document, has been recognized by many textual scholars and even historical linguists who adopt

---

<sup>1</sup> For example the *Brown Corpus* was stored on keypunch cards restricted to uppercase letters (Francis and Kučera 1964), and the *Lancaster–Oslo/Bergen Corpus* corpus in files consisting of lines of 80 American Standard Code for Information Interchange character stored on magnetic tape (Johansson, Leech and Goodluck 1978).

<sup>2</sup> Although the compilers of HC recognized and appreciated the considerable material differences between modern texts and the historical texts contained by the Helsinki Corpus, technological constraints, the strength of the precedent set by the established modern corpora, and the fact that they were basing their work on modern printed editions, meant that their encoding systems maintained a close affinity to those of the earlier Present-Day English corpora.

a *multimodal* approach to the analysis of the text (see e.g. Grazia and Stallybrass 1993, Caie 2003, Suhr 2011, Tyrkkö, Marttila and Suhr 2013, and Hiippala 2013), there have so far been relatively few attempts to integrate these features into historical corpora and consequently into linguistic and textual analyses of historical texts.

This is quite unfortunate and also somewhat ironic, since the analytic representation of such multimodal features in computer-readable form has been extensively studied in a closely related field of study, commonly known as *digital humanities*, focusing on the use of digital methods for studying traditional humanities objects. The irony in the situation stems from the fact that corpus linguistics and linguistic computing were intimately connected with the birth of digital humanities in the 1980s, but for unknown reasons drifted apart over the next two decades. For much of that period, this meant that scholars within one discipline could be completely unaware of the recent developments or even of the established practices of the other. Fortunately, this rift seems to be slowly pulling together, with more and more younger scholars situating themselves equally within both traditions and even being frequently surprised by the existence of such a rift (see e.g. Froehlich 2013 and Hettel 2013). Within the digital humanities community, which has been much more closely associated with literary scholarship and the (digital) editing of literary works than with historical linguistics, the past two decades have seen considerable advances in the use of the digital medium for the representation of textual documents, centred around the development of the *Text Encoding Initiative (TEI) Guidelines for Electronic Text Encoding and Interchange* (TEI Consortium 2014). These guidelines have been developed continuously since their first version published in 1987, and also form the basis for the annotation used for the present edition (see chapter 11).<sup>3</sup> While most early scholarly digital editions and text archives were literary in nature, there is also a tradition of digital documentary editions (see section 4.2), and recent years have seen even the first explicitly linguistic editions of historical material (see subsection 4.3.4).

Like the concept of digital humanities, which is here understood to refer to all research in the humanities that would not be possible without computers and the digital medium, the concept of the *digital edition*, also known as the “hypertext edition, the hypermedia edition, the multimedia edition, the computer edition, [...] and the electronic edition” (Vanhoutte 2006: 161) has no generally accepted definition and is often used without any kind of definition at all. Often it is seen only as an alternative presentation format to the printed book, being essentially similar to a printed edition apart from being displayed on the computer screen and perhaps furnished with some additional features such as hyperlinks from the text to editorial notes or from a table of contents to individual documents. For example the requirements posed by the Association for Documentary Editing Committee on Electronic Standards (ADE-CES) for an electronic edition are no different from those of a printed scholarly edition:

The ADE-CES defines an electronic edition as primary source material prepared with 1) rigorous attention to the text, 2) explanatory annotation and 3) an explanation of the editorial practices used on the texts.

<sup>3</sup> A brief overview of the *TEI Guidelines* is provided in section 5.7 and the full documentation or the current version of the Guidelines is available at <<http://www.tei-c.org/Guidelines/>>.

(ADE Committee 2002)

While these are naturally solid recommendations for any scholarly edition, they do not imply anything beyond a digitized version of a traditional printed edition. In line with the definition of digital humanities above, this thesis argues for a more rigorous definition for digital editions, requiring them to enable significant new possibilities over a printed edition in order to be considered ‘properly digital’ editions. This thesis thus shares the view of those scholars (discussed in chapter 5) who see the digital edition not as a mere representation but as *an analytical descriptive model* of selected aspects of the original document, classifying and simplifying the analogue chaos of the original manuscript document into discrete digital categories, allowing us to gain better analytical purchase on the features represented by the model.<sup>4</sup>

The problem with the majority of existing digital editions fulfilling this definition is that they are packaged into products—whether commercial or not—accessible only through a specific user interface, most often not designed with corpus-linguistic inquiry in mind. This makes it difficult to repurpose them for linguistic research—or any use not envisioned by the designer of the interface—even if the data itself would be useful. Another, perhaps even more problematic result of this ‘productization’ is the fact that it is usually not possible to add further data or metadata to these kinds of editions after they are considered ‘complete’ by their editors or compilers, often in the name of preserving the integrity of the edited text.<sup>5</sup> This kind of a closed architecture makes it difficult to link or integrate further analytical metadata derived from research to the edition itself, effectively preventing it from becoming a nexus for the accumulation of scholarly analysis, annotation and commentary on the edited text. In order to change this unfortunate state of affairs, this thesis will openly advocate the development of digital editions that are “open to incorporating work from everyone who is interested in contributing” without compromising the integrity of the existing editorial content.

One of the core ideas of this thesis is that this kind of openness would help in fostering cross-disciplinary co-operation between digital editors and scholars of various fields interested in the edited material. Considering that the number of people that are interested both in corpus linguistics and manuscript editing is quite limited, this kind of cross-disciplinary co-operation and *division of labour* between scholars from different fields is crucial for the creation of rich and versatile textual resources for the study of our written past. This is not an entirely novel idea. For example Ore (2009) has proposed a similar model for editing, where the separate aspects of editorial work are seen as separate modules which can be performed by different scholars or organizations and do not need to take place all at once. The central idea of this approach is to design editions around the concept of an

<sup>4</sup> An additional benefit of this view of a digital edition is that since it is not tied to any specific presentation but describes the original document in abstract terms, a digital edition can be analysed and visualized using a variety of tools and methods, both ones existing today and ones developed in the future.

<sup>5</sup> Paradoxically, this kind of a closed structure is in fact a threat to the integrity of the edition or corpus, as the only way to add further metadata is to directly edit the original data and possibly lose some of the original data in the process if the encoding system cannot support multiple layers of metadata. Examples of this include the linguistically annotated versions of *The Helsinki Corpus of English Texts*, which sacrificed the annotation of what little paratextual data there was in order to accommodate the addition of linguistic data to the text.

*archive* of editorial material (facsmiles, transcriptions, annotation) which “is what will be carried into the future and what can be used for future editions” (115) and “can be built upon by philologists of the future” (114). This thesis shares the view expressed by Rehbein (2008) that an edition should not be seen as an end in itself but “mainly a resource for future research, [...] a tool for study purposes” (5).

This kind of modular thinking implies an edition that would never be ‘complete’ and would not need to be the result of a singular scholarly process culminating in a final act of publication. As Ore (2004) argues, it would in fact be more useful to organize even the editorial process as a series of stages, each building on the results of the previous one and the results of each being published as they are completed:<sup>6</sup> the production of high-quality facsimile images being followed by a high-quality encoded transcription of each manuscript text, which could then be enriched by further layers of analytical annotation, which in turn may be used to produce a critical edition.<sup>7</sup> This sequential and cumulative model of producing digital research resources hinges on the optimistic view that scholars are willing to build on existing open research resources and to share the product of their own labour on similar terms. This optimism, fostered by the success of the Open Source movement and its derivatives in other fields, is also gaining ground among scholars, who, like Eggert (2009: 79), believe that under favourable conditions this kind of thinking could result in “a productive flow-on effect”:

Let us assume that the quoting scholar has used and benefited from the existing annotation files but realizes they lack a significant dimension. Having harvested, the scholar might also want to contribute this different interpretation of the text, or some local aspects of it, as an additional tagset that may be of use to future users of the site.  
(Eggert 2009: 79)

This kind of open, sequential, and cumulative model has several benefits compared to a more closed and monolithic one. First, it produces usable results more quickly: already the facsimile images serve the valuable purpose of facilitating access, and detailed transcriptions of an individual manuscript texts are not only worth their weight in gold for a historical linguist, but also quite usable by a historian or a literary scholar, especially if the alternative is a critical edition that will not be available for another ten years. Second, it provides more value for the same amount of work, resulting in not only the finished end product, but also an archive of research material in the form of the intermediary facsmiles and transcriptions, which may not only be used to produce different kinds of editions in the future but may also prove to be more useful for certain kinds of research than the ‘finished’ edition.<sup>8</sup>

<sup>6</sup> Since the digital medium radically lowers the cost of the act of publication, there is no real incentive for a single, definitive publication of the entire edition.

<sup>7</sup> It is important to note that not all of these stages need to be performed by the same person or project—and in most cases a division of labour would in fact be the ideal case, with the holding library producing the facsmiles, a team of philologists and palaeographers producing the transcriptions, and scholars with various specializations contributing additional layers of analytical annotation.

<sup>8</sup> Ore (2004: 36) suspects that one reason why editorial projects aim at the production of a monolithic critical edition from scratch is the fact that “scholarly merit has traditionally been linked with the final product” (36) and the creation and development of digital tools—including such interme-

## 1.1 The aim of this thesis

The needs that bring editors to the task of producing a new edition of a work are personal needs. The editor's formal orientation, whatever its basic configuration and however much he qualifies it, is also a personal one. The editor, for his own sake, must fulfil his needs, following his formal orientations with integrity and logic.

(Shillingsburg 1986: 108-109)

As indicated by its title and implied by the above observations on the relationship between digital editing and corpus linguistics, the purpose of this thesis is to develop a well-documented methodology for editing historical documents in the digital medium for the purposes of corpus linguistic enquiry, and to demonstrate this methodology in practice by creating such an edition of a 'family' of six Late Middle English culinary recipe collections commonly known as *Potage Dyvers*. The motivation behind the edition presented in this thesis is thus both historical-philological and corpus-methodological, based on the combination of an interest in the early stages of the historical genre of recipes—and vernacular utilitarian writing in general—and a desire to develop better research resources for the contextualized study of these kinds of historical texts using corpus-linguistic methods on both their textual and *paratextual* content to seek answers to not only strictly linguistic but also more generally *textual* research questions.

Both of these motivations are grounded on a fundamentally *philological* approach to textuality, defined by Robins (2004) as a "*respect for the contingent determinacies of textual phenomena*" and by Wenzel (1986) as an endeavour to understand historical written communication not only in light of its language, but also of all the contextual information that we have access to,<sup>9</sup> or more elaborately:

an appreciative attraction to verbal documents that seeks to understand their meaning, starting with the surface and penetrating to whatever depths are possible, but also alert to the fact that a given text comes from and is shaped by a specific time and place that usually is significantly different from that of the observer. (Wenzel 1986: 12)

---

diate editions—for the service of scholarship has traditionally not been highly regarded by "status conferring enterprises (like tenure committees)", but seen rather as "interesting exercises' akin to, say, generating palindromic verse, or writing a novel without an 'o' in it" (Jensen 2004: 551). Since the role of scholarly editing is an ancillary—and necessary—one in the sense that its objective is to produce resources and tools for research in other fields, something is rotten in the state of the discipline if considerations of scholarly merit and prestige are allowed to divert us from the task of maximising the usefulness of our editions.

<sup>9</sup> According to Wenzel (1986: 12), these include its sources, political and social history, biography of the writer, the socioeconomic conditions of book production, the conditions and processes of manuscript copying, its religious and cultural environment and its intended use and audience.

Although the table of contents might lead one to think otherwise, this thesis is nevertheless first and foremost an *edition*. While the digital nature of the ‘edition itself’—a thorny concept in itself as will become apparent later—has forced its relegation to an appendix located on an enclosed CD-ROM—it nevertheless forms the core around which the rest of this thesis is organized. The printed part of this thesis is intended to justify, document and explain the edition, essentially serving as an—admittedly grossly overweight—editorial essay intended to answer the criticism of Blake (1998) that “it is rare to find any theoretical discussion of editing in editions” (63). But while intended primarily as an edition, it is not intended to be *just* an edition, but also an argument for the usefulness of such an edition and a template for the creation of other editions like it. While it could be accused of indecision between whether to be an edition or a theoretical treatise *about* an edition, I argue that in order to be effective as one, it also has to contain the other: arguing and making demands for a specific kind of edition without accepting the challenge of actually producing such an edition would make the arguments much less convincing and more importantly, less useful, whereas producing a relatively unconventional edition without justifying and documenting its principles and practices would not only be bad editorial practice, but would also be very unlikely to raise constructive discussion on the further development of such editions. Much like for the EpiDoc Aphrodisias Project (EPAPP)<sup>10</sup> described by Roueché (2009), one important function for the present edition is to serve as a demonstration of the suggested annotation methodology and to encourage—if not its adoption by others, at least constructive criticism and discussion that might lead to a more widely acceptable version of a framework for editing historical documents for corpus linguistics.

As indicated by the above discussion, the central issues that are seen as problematic in terms of current historical corpus linguistic practice and need to be solved by any framework for producing corpus linguistic editions include:

- 1) how to analytically model historical documents in terms of both their textual and paratextual content on a level of abstraction useful to a variety of corpus-based linguistic, textual and bibliographical analyses without hindering other potential uses of the edition;
- 2) how to define, encode and present the edition in a way that allows it to be analysed and displayed using a variety of tools and/or converted to a variety of secondary storage or presentation formats;
- 3) how to enable the use of corpus-linguistic analysis methods developed for modern languages for historical texts with highly variable orthography while simultaneously preserving and allowing the analysis of the full range of variation in the original document;
- 4) how to conceptually separate the different levels of interpretation involved in 1) the description of the physical and visual features of the manuscript in analytical terms and 2) the second-degree analysis of what these features mean on a variety of pragmatic and semantic levels;
- 5) how to allow the edition to be perpetually expanded and augmented both with new types of analytical metadata and new parallel texts, not only by the original editor, but also by users and third parties, while simultaneously

---

<sup>10</sup> <<http://www.epapp.kcl.ac.uk/>>



maintaining the integrity of the original data.

This thesis will aim to propose solutions for all these issues, based on the creative application and combination of theoretical ideas and practical solutions proposed and tested by earlier scholarship, making as much use of established standards and practices as possible, both to avoid reinventing the wheel and to maximize compatibility with other existing and future tools and resources.

## 1.2 Historical recipes as editorial material

Cookery books are practical, but can be the starting-point of studies relating to topics far more diverse than merely food preparation. It may seem strange, initially, that cookery books, old and new, can form an important collection within an university library. Recipes are not generally perceived as a source of academic study: can they have the same intellectual, theological, historical or literary values of material in other collections? But everyone must eat, and the procurement, preparation and presentation of food, as it was done over the centuries and continues today, must be of interest to all and is part of the social and economic life of any society. Cookery books can therefore provide source material for a range of disciplines. (White 2004b: 13)

Digital editions, much like printed ones, have focused predominantly on major literary works, essentially replicating the work done by earlier print editors (Leslie 1993: 49), or on the documentary heritage of significant persons or institutions. While medieval recipes and other practical writings—including some of the manuscripts edited here—were edited already in the 19<sup>th</sup> century, these kinds of writings have only started to attract attention since the late 1970s (Keiser 1998c: 109). In their *Guide to Editing Middle English*, McCarren and Moffat (1998) gave two very noteworthy reasons for including chapters dealing with the editing of practical non-literary texts: 1) “much of unedited Middle English falls within this realm”, and 2) “unlike the popular literary texts, these technical works, once they are edited, are rarely re-edited” (McCarren and Moffat 1998: vi).

From this perspective, Middle English culinary recipes have fared relatively well, with a number of both the major collections and minor individual manuscripts having been edited from the 1980s to the present day.<sup>11</sup> Already in 1998, Hieatt expressed her satisfaction in the fact that the field of culinary history “is now being taken seriously by historians (and even some philologists - although those of us in the latter category seem to be an endangered species nowadays), and much more help is available today than was the case only a few years ago” (Hieatt 1998a: 139-40).<sup>12</sup> In addition to being edited, medieval recipes—both culinary and otherwise—have also received scholarly attention as a genre and a text

<sup>11</sup> Although it has to be noted that the vast majority of these were the products of the pioneering work of a single scholar, Dr. Constance B. Hieatt, together with her various collaborators, and her passing in 2011 unfortunately marks the end of an era in the editing and study of Middle English recipes and leaves younger scholars with the challenging task of keeping her editorial legacy alive.

<sup>12</sup> She specifically mentions the *Répertoire des manuscrits médiévaux contenant des recettes culinaires* as an example, but also points out that it is by no means definitive (Hieatt 1998a: 140).

type, especially with regard to their textual organization. While a detailed analysis of the textual organization of individual recipes and the textual features of the recipe genre is not possible within the scope of this thesis, a summary of earlier research on these topics is presented in chapter 8.

The source material for the present edition is constituted by six 15<sup>th</sup>-century manuscript collections of culinary recipes which are closely related to each other and were considered to form a ‘family’ by Hieatt (1992: 21), who named this family *Potage Dyvers* (PD), based on a heading in one of its manuscript versions.<sup>13</sup> These six collections, of which three reside in the British Library in London (MSS Additional 5467, Harley 279 and Harley 4016), two in the Bodleian Library in Oxford (MSS Ashmole 1439 and Douce 55) and one in Durham University Library in Durham (MS Cosin V.iii.11), contain from 169 to 269 recipes and are related to each other in varying degrees in terms of shared recipes and their organization within the collection. The physical features of these manuscripts—which vary considerably—as well as their intellectual content, are described in detail in chapter 9, while the structural relationships between them are analysed in chapter 13.

Although the *Potage Dyvers* family of collections is one of the largest surviving collections of Middle English culinary recipes, it has received relatively little scholarly attention. While four of the six manuscript versions were edited in 1888 by Thomas Austin for the Early English Text Society (EETS), this critical edition is unfortunately not very well suited for linguistic study, as it collates the four manuscript texts into two synthetic versions—combining MS Harley 279 with MS Ashmole 1439 and MS Harley 4016 with MS Douce 55.<sup>14</sup> The fifth and sixth members of the family were discovered only recently, MS Additional 5467 having been described in 2004 by Hieatt, and MS Cosin V.iii.11 in 2006 by myself, when I edited it for my MA thesis (see Marttila 2006).

The fact that there are no earlier digital editions of culinary recipe collections is slightly surprising, considering that as composite texts consisting of short structurally independent units frequently occurring in multiple manuscript versions they are especially well-suited for editing in the digital medium with its capability for explicit hyperlinking and convenient navigation.<sup>15</sup> Even apart from any personal preferences and special interests, Middle English culinary recipes can be argued to constitute an excellent test case for a corpus-linguistically oriented digital edition for several reasons:

- 1) recipes represent a relatively rare utilitarian form of language use which is “often blunt and straightforward, but contains the rhythm and directness of everyday speech rather than the self-conscious language of literary com-

<sup>13</sup> Even though it actually refers to one of the subsections in that particular manuscript, this name has been adopted in this thesis as a convenient label for referring to this family as a whole.

<sup>14</sup> Although Hieatt (1998a: 133), who has also compared Austin’s edition against the manuscripts, notes that Thomas Austin was an excellent editor—at least considerable better than some other early editors of culinary materials—she does admit that “he did misread a word once in a while” (Hieatt 1998a: 134). Furthermore, his edition is also problematic in making various kinds of silent emendations to the text and providing little information on its editorial principles and practices.

<sup>15</sup> *The Malaga Corpus of Late Middle English Scientific Prose* does contain several manuscript collections of medical recipes, but does not contain multiple parallel versions of a single text that would benefit from hyperlinking. It also differs from the present edition by relying on facsimile images instead of descriptive annotation for the representation of the paratextual aspects of the original document and does not allow the user direct access to the annotated transcriptions.

- position”, which makes them especially interesting in terms of historical linguistic study (White 2004b: 13; Claridge 2008: 247);
- 2) as a genre, recipes have an unusually long history during which their basic function—to provide instruction for the preparation of food—has remained unchanged although their situational pragmatic functions have most likely changed significantly, making them a good subject for a long diachronic corpus (cf. Claridge 2008: 247);
  - 3) as structurally complex and composite texts or *discourse colonies* (see chapter 2) surviving in multiple structurally differing versions, recipe collections not only benefit from the linking capabilities of the digital medium but also make for an ontologically challenging text to encode and analyse, testing the flexibility of the encoding and annotation system.

Furthermore, an edition of a family of related recipe collections like the *PD* which contains transcriptions of several contemporary witnesses of the same text is especially useful for the study of certain types of linguistic questions, as it avoids the disadvantages of historical corpora like the Helsinki Corpus which draw “on texts of different genres written at different times”, making it difficult to tell whether changes are caused by variables like genre instead of genuine developments in the language (Blake 2000: 39). Despite its especial suitability for linguistic study, the edition forming the core of this thesis is not intended as a mere methodological demonstration, but as a real contribution to the study of early English culinary recipes. In the future, the edition is intended to form a foundation for a long diachronic corpus of recipe texts from the Middle Ages to the 19<sup>th</sup> century—conceived as a long-term postdoctoral project—which will allow the study of the diachronic development of the recipe genre and its linguistic and textual features throughout the centuries.

### 1.3 The structure of this thesis

The twofold aim of this thesis and the digital nature of the edition itself also pose challenges in terms of its textual organization. The latter issue has been dealt with by including the edition itself—consisting of a number of interlinked TEI XML data files—along with three different digital presentations (a diplomatic transcription, a parallel reading edition, and an interactive HTML edition combining the two) and various other digital resources in a digital Appendix stored on a CD-ROM included with the printed editions and in a separate archive file included with the digital edition of this thesis. The former, namely the concurrent description and contextualization of the edition itself and of the edited texts, has been solved by including two separate background parts, the first contextualising the digital edition in terms of its theoretical background and underlying ideology, and the second providing the historical and cultural information required for the proper contextualization of the family of recipe collections modelled by the digital edition.

These background parts are followed by a third part documenting the edition itself, both in terms of its manuscript sources and its editorial principles and annotation practices—formulated on the basis of the theoretical background presented in part I. The final part again deals with the edited content, consisting of a dialectal analysis of all the manuscript versions to determine their likely areas of

origin and their degree of linguistic standardization, and a text-structural analysis of the six versions in terms of the number of recipes shared between them and the organization of these recipes within the collections in order to shed light on the textual history of the *PD* family. Finally a brief concluding chapter will draw together the principal observations made during the editing and research process regarding both the theoretical and technical framework within which the edition was conceived and what the edition revealed about the edited manuscript texts.

Since both editing and corpus linguistics are focused on the manipulation of *texts*, the first task of any discussion of these fields should be to define what we mean by ‘text’. To this end, chapter 2 will lay the basis for the theoretical framework of the present edition and the editorial approach being developed for corpus linguistic editing by discussing the concept of *text* from an ontological viewpoint and establishing the concept of a multi-level *textual object* that is the object of all editorial activity and textual analysis. After this, chapter 3 will provide an overview of the Anglo-American tradition of editing historical documents and examine the relationships of its two main branches, *critical editing* and *documentary editing*, to 1) the nature of medieval textuality, 2) the possibilities offered by the digital medium, and 3) the requirements of linguistic research, characterising the editorial approach taken in this thesis in relation to these two traditions. The third chapter of the theoretical background part, chapter 4 first describes in detail and then relates to each other the theoretical and practical requirements of historical corpus linguistics, and the theoretical and practical possibilities that the digital medium provides for fulfilling them. Based on this discussion, chapter 5 proceeds to describe an ontology of the digital edition that will fulfil the requirements of linguistic research and help us solve the issues mentioned above.

The historical background part of the thesis that provides the necessary background information for understanding the edited content is divided into three chapters. In order to contextualize the dialectal analysis presented later on in chapter 12, to provide a basis for the discussion of the potential producers and users of written recipe collections in chapter 8, and to provide those users of the edition unfamiliar with the period with the necessary context for the interpretation of the language of the recipes, chapter 6 contains a brief description of the linguistic situation of 15<sup>th</sup>-century England in terms of the different languages in use and their functions, the gradual establishment of a linguistic standard over this period, and the level of literacy. Chapter 7 on the other hand provides a brief description of the sociocultural and situational context in which culinary recipes were produced and used, providing the user of the edition with the necessary context for the interpretation of their semantic content. The final chapter of the historical background part, chapter 8, contains a brief account of the tradition of culinary writing in late-medieval England, characterizes the recipe collection as a composite textual object or *discourse colony*, and provides an overview of the textual and functional characteristics of recipes as a *text type* and a *genre*, followed by a brief discussion of the potential producers and users of recipes collections based on earlier historical research.

The third part of the thesis again focuses on the edition itself. Chapter 9 focuses on the *Potage Dyvers* family of recipe collections, establishing the identity of the family, providing a detailed codicological description of the six different manuscript versions, and situating it within the English and continental recipe

traditions by charting its relationship to various other families of collections or individual collections with which it shares some of its material. Chapter 10 contains an account of the principles followed in creating the edition and the various decisions and editorial judgements that have been made in preparing it, relating them to the theoretical discussions of chapter 4. Building on the previous chapter, chapter 11 provides a detailed description of the specific set of eXtensible Markup Language (XML) elements, attributes and attribute values that are used to model the different aspects of the original document.<sup>16</sup> Together, these two chapters documenting the principles and practices of the edition are also intended to form the basis for the development of a more normative set of guidelines for creating digital editions of historical documents for the purposes of corpus linguistics.

While the claim of the present edition to the status of a ‘corpus-linguistic edition’ would best be substantiated by the inclusion of corpus linguistic analyses of the text, the sad fact is that most types of traditionally corpus-linguistic inquiry are not especially interesting or relevant when performed on what amounts to a single work, even if represented by six different versions. For this reason, part IV focuses on the analysis of issues that are particularly interesting in the context of the *PD* family. These include 1) the detailed analysis of those linguistic and orthographic features presented as dialectally significant by the *Linguistic Atlas of Late Middle English (LALME)* to establish the most likely dialectal origins and level of standardization for the six manuscript versions (chapter 12), and 2) the macro-level structural analysis of the six collections in terms of the individual recipes contained by them and their ordering within the collection in order to establish the relationships between the six versions in terms of their textual pedigrees (chapter 13). The properly corpus-linguistic analysis of Middle English recipes as a genre will be deferred to a later date following the completion of a more comprehensive corpus of Middle English recipes edited following the guidelines presented here.

Since the edition that forms the core of this thesis exists natively in the digital medium and is thus impossible to represent in printed form, this thesis also contains a total of seven digital appendices containing not only the edition itself as a digital *data archive* but also a formal definition of its syntax, and a variety of editorial outputs based on that data archive together with the eXtensible Stylesheet Language Transformations (XSLT) transformations used to generate them, along with the raw data used for the analyses presented in chapters 12 and 13. The fifth part of this printed edition thus consists of a series of appendices describing these different kinds of digital data and providing information on how to find and access them on the accompanying CD-ROM<sup>17</sup>, which contains the actual content of each of the appendices.

<sup>16</sup> The annotation system described in this chapter is based on the *TEI Guidelines for Electronic Text Encoding and Interchange* and essentially defines a customized, more narrowly specified and somewhat extended subset of the Guidelines tailored to the requirements outlined in the preceding chapters.

<sup>17</sup> In the digital edition of this thesis, the contents of the CD-ROM are contained in a .zip file distributed with the thesis.



## **Part I**

# **Theoretical background**





## Chapter 2

# Conceptualising the text

Indeed at first glance the word *text* is so unambiguous that no further effort to define it seems necessary. Its everyday usage appears as unproblematic as the words *bread*, *house*, and *dog*. [...] Hence, for a long time editors had no motivation for a more far-reaching discussion of the central concept of their occupation. (Martens 1995: 209)

It is not only editors that often take the concept of *text* for granted: historical corpus linguists, text encoders and literary scholars are just as likely to use this word as a rather loose general term, assuming it to be sufficiently well-defined by the conventions of their discipline. This may be true when communicating strictly within a single scholarly discourse community, but once editors start talking to historical linguists or literary scholars to text encoders, the lack of definitions is bound to cause confusion. As the express purpose of this thesis is to build bridges between the disciplines of historical corpus linguistics, manuscript editing and electronic text encoding, it is necessary to begin the discussion of these matters with some kind of conceptual clarification.

Considering the interdisciplinary nature of the present discussion, it is important to define the concept of text—and its relatives—in a way that is sufficiently relevant from both linguistic and editorial point of view and helps to relate these two disciplines to each other. While both text linguistics and textual scholarship are ostensibly concerned with the same object—the text—this has in fact not been the case for the most part. Textual scholarship, the theoretical backbone of text editing, has traditionally been based on a very different concept of text than linguistics: while the former sees the text as something to be “established” by an editor through the judicious emendation of various source documents—in other words, a *product* of scholarly activity—the latter sees it as a “naturally occurring manifestation of language” (Beaugrande and Dressler 1981: 63), i.e. the *object* or *source material* of scholarly activity.

This in itself would not be a problem, as it makes the two disciplines not so much incompatible as complementary. The problem, however, lies in the diametrically opposed perceptions of the role and situation of the text that these viewpoints entail. Whereas a textual critic sees the text as an embodiment of and a

witness to an abstract literary *work*, a linguist sees it as an instance of language use encoded in a physical *document*. This different perception of the ontological situation of the text and of its significance have resulted in irreconcilable differences in both the expectations made of historical texts and the procedures they have subsequently been subjected to, leading editors to produce ‘texts’ which do not correspond to what a linguist considers a valid historical ‘text’.

## 2.1 Documents, texts, versions and works

This ‘multi-level’ nature of textual objects, alluded to above, has been noted by many textual scholars and discussed in the most systematic way by Shillingsburg (1986), who defined the concepts of *document*, *text*, *version* and *work* to describe the different ontological levels of the textual object. In Honkapohja, Kaislaniemi and Marttila (2009: 453) we introduced a modified and simplified version<sup>1</sup> of this multi-level conceptualization, consisting of the levels of the *artefact*<sup>2</sup>, *text* and the additional concept of *context*, based on the notion of texts as cultural products espoused by Jerome McGann (see e.g. McGann 1991).

Since different types of editions can be seen to focus on different levels of textual conceptualization—critical editions focusing on the *work* and diplomatic or documentary editions on the *document*—and because the relations between concepts and conceptual analyses are “crucial ingredients” of both editorial and bibliographical theory (Dahlström 2009: 29–30), the present section will briefly define the different levels of the textual object as understood in this thesis. The conceptualization of the textual object used here is based both on the conceptualization of Shillingsburg (1986) mentioned above and on the hierarchic ontology of bibliographical entities outlined in the *Functional Requirements for Bibliographic Records* (FRBR: 17–25) which differentiates between *work*, *expression*, *manifestation* and *item*. The concept of *context* is here seen from a pragmalinguistic point of view as a multi-layered phenomenon in itself, as will be described in section 2.2 below.

### 2.1.1 Document

A document consists of the physical material, paper and ink, bearing the configuration of signs that represent a text. Documents have material existence. Each new copy of a text, whether accurate or inaccurate, is a new document. (Shillingsburg 1986: 51)

This definition, also quoted by Machan (1994: 7), posits the *document* as the ‘lowest’ level of a textual object in terms of abstraction, objectively existing in the

<sup>1</sup> This simplified conceptual model of the textual object was not intended as a comprehensive description of textual phenomena, but merely as an initial frame of reference to illustrate the interrelationships of the different types of features that should be encoded in a corpus-linguistic digital edition that we envisioned in the article.

<sup>2</sup> The term *artefact* was used as a replacement for Shillingsburg’s *document* in order to avoid confusion with the computing sense of the term.

material world as ‘marks on the page’. As such, the *document* is also the *pre-textual* level of the textual object—it does not so much ‘contain’ a text as it provides a blueprint for a text, to be produced as “the result of a socialized reader’s engagement with the document” (Eggert 1994: 2). From the point of view of the producer of the document, it can be seen as an *encoding* of linguistic entities (words, letters, punctuation etc.) using conventionally agreed-upon symbols inscribed in significant configurations in some medium.

What is here called a *document* corresponds roughly to the *item*, which is defined by Madison et al. (FRBR: 24-5) from a more bibliographic viewpoint as a physical instance of a *manifestation*. The concept of *manifestation*, which does not really have a counterpart in Shillingsburg’s ontology, refers collectively to all items or documents “that bear the same characteristics, in respect to both intellectual content and physical form” (FRBR: 21). Being clearly a function of printed texts, this concept is not really applicable to a manuscript context where each individual item by definition represents a different manifestation.

Although the ontological primacy of documents as “text carriers and text witnesses” and “the material substratum of textual transmission”, without which written texts cannot exist is usually implicitly recognized, they are nevertheless often relegated to a secondary position by editors, whose cultural and editorial interest has traditionally been directed primarily at the text and the work (Gabler 2007: 197). However, this thesis takes the view represented by Gabler (2007: 199–201) and Zeller (1995a), who have stressed the importance of the document as the factual basis of all textual scholarship:

The only objective thing is the unique original manuscript itself; it may not be replaced by an equivalent, and in the strict sense may not be reproduced even by color photography. The material manuscript itself, not the text of the manuscript, is the record. The manuscript requires interpretation, however, and the result of the interpretation is the text.  
(Zeller 1995a: 43)

Since the process of seeing, analysing and interpreting always moves from the *document* through writing to *text*, we as editors should be “putting the horse of the document properly before the cart of the eventually emerging text” (Gabler 2007: 201), and see the text “fundamentally as a function of the document” and not the other way around, since “it is documents that we have, and documents only” (199).

The established distinction between ‘original’ and ‘copy’ entailed by the mass-production of copies by printing, has often resulted in the original and unique nature of every manuscript being overlooked by editorial theories and practices developed for the editing of printed literary works:

For modern book owners, our individual possessions are copies: the original is the author’s typescript or the publishers plates. Medievals treated each faithful transmission of a text as a genuine original.  
(Toon 1991: 83)

This veneration of the (authorial) ‘original’ document as the most accurate embodiment of the *work*, together with the fact that such an item rarely survives,

is what lies at the heart of traditional textual criticism, as will be seen in section 3.1.

In recent years, more attention has been paid to the palaeographical and codicological features of documents and their role in the construction of texts. This has also led to arguments for a more ‘document-based’ approach to manuscript studies, recognising the fact that the information contained in manuscripts as documents is not limited to the purely textual:

In any branch of manuscript studies (editing, codicology, palaeography, art history, history) the first level of enquiry always is (or should be) the document, the physical support that lies in front of the scholar’s eyes. The fact that the text was transmitted to us by means of a specific physical object which has been organised in a certain way and preserved in one place or another has all sorts of consequences in the way we understand and receive that text.

(Pierazzo and Stokes 2010: 398)

In the current discussion, a textual *document* is understood to refer to ‘a physical object (or in some cases a group of related objects) of some kind, bearing graphical signs that can be interpreted linguistically to produce a *text*, arranged in specific ways that reflect their status and relationships with each other, along with various kinds of non-linguistic or metalinguistic signs’.

### 2.1.2 Text

A text is the actual order of words and punctuation as contained in any one physical form, such as manuscript, proof, or book. A text is the product of the author’s, or the author-and others’ physical activity in the attempt to store in tangible form the version the author currently intends. And yet a text (the *order* of words and punctuation) has no substantial or material existence, since it is not restricted by time and space.

(Shillingsburg 1986: 49-50)

In quoting Shillingsburg’s definition, Machan (1994: 7) furthermore reminds us that neither is a text “equatable with a specific, substantial existence”, as the same sequence of words and punctuation may in fact appear in several documents. As Eggert (2009: 73) has observed, neither do texts have “an unproblematic objective existence”: they are not “self-identical”, but a product of an interaction between a document and an interpretive subject. This means that the relationship between document and text is not a closed one-to-one relationship but rather an open many-to-many relationship in at least two senses. First of all, a single physical document like a medieval miscellany or a commonplace book can sequentially encode a number of independent texts,<sup>3</sup> and secondly, a given document (or a part of one) can give rise to a multitude of different, concurrent texts, depending on

<sup>3</sup> A single text is here understood in its linguistic sense as defined by Werlich (1983: 23-6) and discussed in more detail below.

the interpretive activity applied to it, and conversely, with suitable interpretive parameters, the same text—an identical sequence of words and punctuation—can be produced from several documents.

This fluid and processual nature of text and its ontological distinctness from the physical document has not always been acknowledged by editors, and it is easy to see that an object-oriented—and thus objective—concept of text can seem appealing from an editorial point of view. However, as Martens (1995) has pointed out, “the editor’s activity immediately reveals that limiting the concept of text to the sign carrier, or to the material artifact, is not feasible”, as “the linguistic object fixed in writing [...] is only identifiable as text if the dimension of meaning is perceived (that is, interpreted) at the same time” (215-6). In other words, the “configuration of signs” inscribed in the physical document only becomes text, when it is read and thus imbued with some kind of linguistic or pragmatic significance. In addition to being erroneously identified with the *document*, the concept of text has also been used—mainly by literary critics—“as more or less synonymous with *work*” (Williams and Abbott 1999: 68). This confusion is most likely a result of the editorial desire for a single ‘correct’ textual representation of a literary work that has strongly characterized the field of textual criticism and editorial theory (see section 3.1) and created an illusion of works having a stable textual identity.

In addition to the ‘external’ or ‘depth’ ontology relating the *text* to its material manifestations and ideational abstractions, texts can also be seen to have an internal structural ontology, whose particular ontological question is: “what is text?” (Renear 2001: 34.) This question can be understood both in a qualitative and a quantitative sense, the latter rephrasing it as “what is *a* text?”

### Textual ontology

In qualitative terms, Renear (2001: 35) lists a number of ways in which a text may be viewed, but only two of them are compatible with our definition of the text as an organized collection of linguistic elements, namely the text as a string of characters, and the text as an Ordered Hierarchy of Content Objects (OHCO).<sup>4</sup> In both linguistics and textual criticism, the traditional view has seen the text as a linear, variously segmented sequence of characters. Upon closer inspection, this view is quite limiting as it reduces the text into a single unbroken line or a “data stream” (Sinclair 2004: 4) with no possibility to account for parallelism or divergence, such as is represented by marginal notes or commentaries, or by alternative readings of ambiguous signs found in the document (Curzan and Palmer 2006: 26). The other option suggested by Renear (2001), namely seeing the text as an ordered hierarchy of content objects has been found to be the most effective way of conceptualising texts “by engineers, publishers, and text encoders as general approach to working with text” (Renear 2001: 35). This is the ontology underlying both syntactically parsed treebank corpora and the XML family of markup languages:

In this account text is a particular sort of structure of objects such as chapters, sections, paragraphs, sentences, acts, scenes, speeches, stanzas, metrical lines, and so on. It is hierarchical because these ob-

<sup>4</sup> The other two options mentioned by Renear, a combination of shapes and a hierarchy of layout objects, are clearly topological conceptualizations of the document level.

jects nest inside one another like Chinese boxes, and ordered because there is a linear order. The objects mentioned are called ‘content objects’ because they are in some sense units of sense and meaning – such things, on this view, are the stuff texts are made of.

(Renear 2001: 35)

Because it maintains the same linear structure while adding to it a second dimension, it can be viewed as an extension rather than a replacement of the linear character string ontology. While the OHCO model has been found to be an effective conceptualization for the hierarchical structure of text as a series of chapters, consisting of headings and paragraphs, in turn consisting of series of sentences which are made up of clauses and phrases, which are in turn made up of words, it has a single major shortcoming: it assumes the textual objects to “nest within each other without overlap” (Pierazzo and Stokes 2010: 419), which means that an object cannot be contained by two other objects of the same nesting level (i.e. if one of them is not contained by the other). This means that it does not allow for things like a syntactic model of a word functioning simultaneously in two different clauses or phrases. In its traditional form, its adherence to the linear order of the objects means that as a conceptual structure it cannot represent parallelism or alternatives in the textual flow.<sup>5</sup> It should also be noted, as Pierazzo and Stokes (2010: 420) point out, that the OHCO model is still only two-dimensional and can only model a single hierarchy—usually the linguistic text level of the textual object—which means that it does not capture the connection between the text and the document levels of the textual object.

As a form of *digital documentary edition* (see sections 3.2 and 4.2), the present edition focuses on the representation or modelling of the document and the different texts that can be produced from it instead of the establishment of a single authoritative text. Thus the textual object represented by the edition can be seen to contain multiple potential texts and conceptualized as a *multidimensional relational network of content objects*, where the individual content objects or textual elements are related to each other in a variety of ways, indicated by a combination of the visual cues in the document (see subsection 2.1.5 below) and the linguistically, situationally and culturally determined conventions that guide their interpretation<sup>6</sup>. A four-dimensional example of such a conceptualization would be to see each word as a textual unit connected longitudinally to the words preceding and following it in the linear stream (linear order within a larger unit), latitudinally to a marginal gloss characterising it in some way (parallelism between two separate larger units), altitudinally to the larger unit made up by all the words belonging to the same sentence (hierarchy between a larger and a contained smaller unit), and chronologically to a word that was erased and replaced by the current

<sup>5</sup> This should not be confused with the use of local instances of the OHCO model with suitable semantic definitions for representing alternation or parallelism, which is a common practice in text encoding and will also be used in this edition (see e.g. subsection 11.4.2 and section 11.7).

<sup>6</sup> For example in the context of medieval European manuscripts, a word placed immediately to the right of another word on the same line within the writing block would be interpreted to follow the other word in the textual stream, while a word written above another word that has been struck out would be interpreted to chronologically supersede it, and a word or phrase placed in the margin at the same height as an underlined or otherwise marked word within the writing block would most likely be interpreted to provide a parallel expression or an explanation of the other word.

word (alternation between two mutually exclusive units). By omitting or collapsing together some of the dimensions of this multidimensional network of textual units, a variety of lower-order representations can be generated, their ontologies depending on the number of dimensions preserved.<sup>7</sup>

### Delimiting a text

The quantitative definition of *text* is especially important in the current context, because it also serves as the basis for the definition of a *work*. As was already mentioned in passing, a *text* is here understood in a text-linguistic sense as “an extended structure of syntactic units such as words, groups, and clauses and textual units that is marked by both *coherence* among the elements and *completion*” (Werlich 1983: 23). *Coherence*, according to Werlich “is created in linguistic communication whenever the encoding communicant (*encoder*) uses some linguistic unit (usually a *group* of words or a *sentence*) as a *text base unit* with a theme and then expands this unit in linear progression in conventionally ordered and completed sequences of linguistic units” (23).<sup>8</sup> Put more plainly, a text is a piece of discourse that introduces a topic and makes statements about it in a coherent manner.

*Completion*, on the other hand, “is created in linguistic communication whenever the encoder introduces signals which indicate both the beginning and the end of one or more of the sequences that have established coherence” (24). As Werlich points out, the *initiation* and *termination* of texts “depends on the adult communicants’ acquired *textual presuppositions* about a well-formed text” and their ability to recognize the sequences conventionally used to signal them, which are of course historically and culturally determined.<sup>9</sup>

### Discourse colonies

While the above characterization of text can be considered an accurate description of the kinds of texts historical linguists have traditionally concentrated on, consisting of segments of *connected* discourse, a large proportion of the material contained in medieval manuscripts—“proverbs, law codes, calendars, recipe collections, and mnemonic jingles” (Carroll 2003: 137)—consists of short segments of *unconnected* discourse that do not comfortably fit the above definition of a text even though they are in other respects considered as ‘texts’. For describing such composite entities, made up of small independent texts combined not by traditional coherence but by a common semantic and structural frame, Hoey (1986, 2001) has introduced the concept of a *discourse colony*. In describing the structure of a discourse colony in relation to normal texts, Hoey metaphorically likens it to a *beehive*:

<sup>7</sup> For example a pure OHCO representation would preserve two dimensions (linear order and hierarchical structure) while a plain text would preserve only the linear order.

<sup>8</sup> These sequences can be either *functional* or *topical* in nature (Werlich 1983: 30), which essentially means that the constituent parts of a *text* are linked together both by functional ties such as consistency of tense and number, sequence signals, deixis, etc. and by topical ties, i.e. the recurrence of semantic components throughout the text.

<sup>9</sup> Fortunately, these textual boundaries are usually also signalled on the level of the document by various visual elements (see subsection 2.1.5), allowing us to distinguish them even when our presuppositions about the linguistic signals differ from the medieval ones.

‘Mainstream’ texts are like people. They are made up of interconnecting parts and typically have a single author. The interconnecting parts combine to form conventional units (like paragraphs, sections or chapters) and the random excision of one of these will frequently impair and sometimes render unintelligible the text. Like people, if such texts are jumbled up, they die (though, unlike Humpty, they can in some circumstances be put together again). [...] What distinguishes the beehive from the human being is the fact that the former’s organisation does not depend on its parts being connected in one and only one way: the bees enter the hive in no order. This property of the hive is sufficiently distinctive for it to serve as a working definition of the discourse colony: let us say that *a colony is a discourse whose component parts do not derive their meaning from the sequence in which they are placed*. If the parts are jumbled, the utility may be affected but the meaning remains the same. (Hoey 2001: 74-75)

In addition to the main defining feature of a discourse colony, Hoey (2001: 77–87) has identified eight other properties of a prototypical discourse colony. While not all discourse colonies manifest all of them—dictionaries and encyclopedias being the only genres found by Hoey to exemplify all nine—Hoey considers there to be “sufficient similarity between types of colony to suggest that the category is not a valueless fiction” (77).<sup>10</sup> The nine properties identified by Hoey (1986: 7–19) and subsequently repeated both by Hoey himself (2001: 88) and other scholars like McDermott and Walsh (1991: 49) and Carroll (2006: 315) are:

- 1) Meaning not derived from sequence
- 2) Adjacent units do not form continuous prose
- 3) There is a framing context
- 4) No single author and/or anon
- 5) One component may be used without referring to the others
- 6) Components can be reprinted or reused in subsequent works
- 7) Components may be added, removed or altered
- 8) Many of the components serve the same function
- 9) Alphabetic, numeric or temporal sequencing

As Hoey (2001: 87) notes, some of these features are more central than others, the first two—of which the second is essentially a corollary of the first—being the ones that are common to all discourse colonies.<sup>11</sup> However, as Hoey (2001: 78) points out, this does not mean that a discourse colony needs to be totally unstructured or that discourse colonies could not contain adjacent elements that can be read as continuous prose, but merely that equally strong connections must be possible also between non-adjacent sections, e.g. through cross-references. Order can also be imposed through hierarchical nesting, one discourse colony functioning as a component of another, in which case the integrity of the sub-colonies must be maintained, even though their components can occur in any order within

<sup>10</sup> Essentially this means that the category of discourse colonies is partially defined through the principle of Wittgensteinian *family resemblance*, although the initial criterion of non-sequentiality can be considered as being shared by all discourse colonies.

<sup>11</sup> Hoey (2001: 87) also observes that the second property could be used as the single defining property just as easily as the first one.



the sub-colony, and the sub-colonies themselves can occur in any order within the main colony (Hoey 2001: 76). While discourse colonies do not rely on the order of elements for meaning, most colonies, “for the sake of utility, make use of some form of arbitrary or non-arbitrary sequence to make selection, reference, and cross-reference possible” (Hoey 2001: 86), although for colonies like newspapers and (modern) cookery books, the division of their components into thematically-based subcolonies, together with indices or lists of contents referring to individual components by page numbers usually eliminates the need for an alphabetical or numerical organization.

Although Hoey (1986: 22-3) characterizes discourse colonies as a *discourse type*, they are more appropriately seen not as a single text or discourse type, but rather a loose set of *discursive features* associated with a number of different text types and genres. The close association of the discourse colony with certain specific genres and registers is indicated by Hoey’s emphasis on the framing context, which he sees to be commonly established by a heading or label identifying the genre or register of the discourse colony (legal act, cookbook, etc.). This association is not surprising if we accept Hoey’s (1986: 22-3) argument that different discourse features have arisen to correspond to different reading strategies, the discourse colony having developed for reference reading, i.e. for locating a specific relevant component—such as a recipe—and for utilising it without regard to those components not relevant to the present purpose (Hoey 2001: 89). Since recipe collections exhibit most of the features described above, the discourse colony is a natural conceptualization for the recipe collection as a whole, and will be used as the basis for the description of the recipe collection as a genre and text type in section 8.2.

### 2.1.3 Version

Most people, even pragmatically minded encoders, would concede that there is no such thing as a single text; any text offers a plurality of readings of its constituents and a plurality of paths along its syntagmatic axis.  
(Haugen 2004: 89)

The problem of the variation and parallelism found in texts, mentioned above, can be solved in two basic ways. Martens (1995) outlines these two alternatives as the two positions in a debate over the nature of text:

- 1) we can either view text as “a unified and closed linguistic object and thereby something fixed” (210), with parallel variants not belonging to the text but being deviations from it and constituting a different text, or
- 2) we can see text “as a complex of all the versions and deviations” that can be produced from the document, with the parallel variants not “constituting different texts, but different *versions*” (210) of the text.

While the second of the options outlined by Martens would intuitively seem the more convenient one, avoiding the fragmentation of the text, the current discussion will in fact opt for a combination or compromise of these options, modelled after Shillingsburg (1986), who uses the term *version* slightly differently from Martens to refer to a version of the *work* instead of a version of the *text*, defining an authorial *version* of a work as “the sequence of words and punctuation the author intended to put in readable form” at a certain point in time,<sup>12</sup> moving it one step of abstraction above the *text*. He notes that while a manuscript of a work is not the version itself, it “contains a fairly accurate representation of it” (48). Modifying this definition slightly, we can postulate *version* as what Martens called *text* in his second scenario, leaving *text* to refer to the closed, unitary linguistic item described in the first scenario.

Thus, a *version* is here seen as an ‘abstract linguistic realization of a work’ (as defined below), and is represented by all of the possible texts that are encoded in a single unique document.<sup>13</sup> This means that it is simultaneously more abstract than the text—corresponding in practical terms to a sequence of regularized lexical items (i.e. words *sans* their specific orthography)—and more closely associated with a single document than the text, as there is a one-to-one relationship between a manuscript document and the version represented by it. Thus we can speak of “the uncorrected text of the manuscript X version of the work”, which is the text produced by ignoring all later corrections made to the document, or “the abbreviated text of the manuscript X version”, referring to a transcription of the document which preserves any signs interpreted as abbreviation markers in the document based on their appearance, with no regard to the textual content they could be interpreted to signify.

The movement from *text* to *version* always involves a movement from multiplicity to singularity, at the cost of loss in detail. The relationship between a *text* and a *version* can be seen as analogous to the superimposition of several similar but not identical line drawings or pages of print on a transparent ground on top of each other: the result is a single image, where the parts that are identical in all the versions are reproduced accurately, and any loci of variation are highlighted by being simultaneously bolder and more blurred. In other words, a *version* is the aggregate sum of all the texts produced by the different reading strategies applied to a single document.

This additional level of the *version* allows us to simultaneously preserve the static, closed notion of text as a “structured, thematically closed unit of linguistic communication” (Martens 1995: 210), while acknowledging the versioning notion of text as dynamic and constantly changing (211), seen “not as an object but as an action” (McGann 1991: 183). Although we thus cannot—as McGann has maintained—talk of ‘the text itself’ of a work as a single unitary utterance, we can define a text that can be justifiably characterized as such. This might well be the

<sup>12</sup> Although Shillingsburg discusses the different authorial versions of the work, the concept and definition applies equally well to scribal versions in a manuscript context, since a scribe can be seen to act as both a reader and an author in copying a text (see subsection 2.3.3 and section 2.4 below).

<sup>13</sup> This definition has been formulated with manuscript documents in mind, which are unique by definition. In the case of printed or electronic documents, which can exist in multiple identical (or quasi-identical) copies, all documents representing the same *manifestation* (as defined by the Madison et al. (FRBR: 21)), i.e. “all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form” are considered to contain the same *version* of the work.

closest we can get to satisfying the readers' "need of an unequivocal text", born out of "apathy, exhaustion, innocence, or complacency" (Pearsall 1985: 93), while still maintaining an awareness of the multiple and variable nature of medieval textuality.

### 2.1.4 Work

A work has no substantial existence. Nor is it a platonic ideal, that is, one fixed ideal form. [...] It is only partially represented by any one given printed or written form of the work. The redundancy of its various printed and written forms gives a sense of unity which helps us to conceive of the range of forms as one work, but its variants suggest the haziness of its outlines. (Shillingsburg 1986: 46)

The concept of *work* is traditionally associated exclusively with literary writing and evokes images of artistic creation by a specific individual, the *author*. This association is clearly expressed by Machan (1994: 6–7), who elaborates on Shillingsburg's definition, describing a work as "the message or experience implied by the authoritative versions of a literary writing" that is "represented more or less well and more or less completely by various physical forms, such as manuscripts, proofs, and books".<sup>14</sup> As a distinctly non-textual entity, the *work* would at first glance seem to have very little relevance for the historical linguist who is, after all, interested specifically in the actual linguistic *forms* used in a given period.

While the *work* may not be interesting for the linguist as an object of study, it can nevertheless be seen as a useful classificatory device similar to *register* or *genre*, only on a more fine-grained level. Following the characterization of a *work* by Searle (2004: 16) as "not some kind of marble monument", but rather "our being led to think in a particular way, with exact and formal guidance", we can define the work, from a communicative point of view, as the *specific communicative function* shared by all the texts representing the work, or the information or impression they try to convey. While admittedly subjective and open to interpretation, this definition of the work has the benefit of avoiding any reference to the printing-era concept of 'author', which has traditionally served as the linchpin for literary definitions of a *work* (Eggert 2004: 171).

<sup>14</sup> While Machan (1994: 7) argues this textual-critical sense of *work* to be fundamentally different from that of contemporary interpretive criticism, "which variously regards the work as process, or cultural artifact, or mutual construction of writer and reader, or infinitely deferred ideal", the concept of work seems to be somewhat vaguely understood even within the text-critical community. For example Eggert (2004: 173) sees the *work* as the result of the "always provisional" act of reading, where "the reader participates in the textual dimension and, in doing so, asks questions, prompted by those practices, about the meaning of the letterforms and the other physical features that he or she encounters", making a direct jump from the physical level of the document to the semantic level of the work, as if "the letterforms and the other physical features" had some kind of inherent "meaning". His characterization is in fact very similar to that provided by Barthes (1979)—arguably representing "contemporary interpretive criticism"—who compares the *text* to the score of a post-serial musical piece, which the reader performs or executes as an active collaborator to produce the *work* (80).

The relationship between a *work* and a *version* can be seen as similar to the relationship between a *version* and a *text* in that they are both one-to-many relationships, the *work* being essentially an abstraction of a group of *versions*, just as a *version* is an abstraction of several *texts*. Thus, while a *version* is always represented by a single unique document (or a series of identical documents), a *work* can be represented by several different documents. In cases where no two documents contain what could be considered the same *work*, the difference between work and version collapses, the work being represented by a single unique version.<sup>15</sup>

Although the relationship between *text*, *version* and *work* as a strictly unidirectional many-to-one relationship, the medieval situation was slightly more complex. As Machan (1994: 162) has observed, medieval textuality was more fluid than what we are used to, the work being less closely tied to a certain text; not only could different versions of a work quite freely rearrange the material and add or omit sections of it, a certain portion of text could also function in an entirely different work, essentially creating a many-to-many relationship between text and work.<sup>16</sup> In the same spirit, longer verse texts were “typically assembled in what might be called a mode of infinite expandability” (Machan 1994: 164), meaning that they could be expanded or contracted at any point by adding or omitting material, resulting in versions that differed greatly in length or extent even though their shared parts might be very similar. This feature is also shared—even to a greater extent—by recipe collections and other discourse colonies, whose component parts can by definition be feely rearranged, omitted or added, as will become obvious in chapter 13.

This, of course makes the question of where to draw the line between different *versions* of the same work and entirely different *works* even more acute than usual: should the six manuscripts edited here be considered to represent six versions of a single work, six different works, or something in between? As Buzzetti and McGann (2006) have pointed out, the work is by nature a fuzzy concept: the various bodies of material that comprise a ‘work’ “orbit around the conceptual thing” that is for convenience named the ‘work’ and “relate to that gravity field in different ways” (70). Considering that the addition and deletion of material is a common way of creating revised versions of a work, and discourse colonies explicitly allow for the addition or removal of functionally equivalent components without changing the identity of the colony, the question is basically one of degree: how much material can be added and/or removed before the resultant document must be seen to constitute a new work? Since the components of a colony are functionally equivalent, and thus equal in ‘weight’, the natural threshold for considering two versions of a discourse colony to represent different works would be the point where both contain more components that are not shared between them than components that are.<sup>17</sup> In reality, the situation is further complicated by the fact that in addition to the *extent* of the material contained by the different versions, also

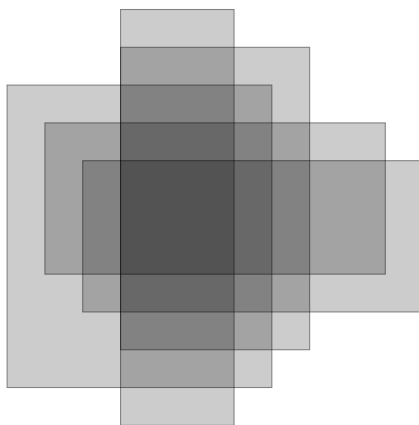
<sup>15</sup> Among literary texts, Machan (1994: 167) sees medieval romances as a genre where “faulty copying, lapse of memory, and conscious alteration by a *disour*” are so frequent and extensive, that even texts ‘copied’ from the same source (or one from the other) constitute independent works, making the distinction between a *work* and a *version* “particularly tenuous”.

<sup>16</sup> Machan uses Lydgate’s *Fall of Princes* as an example, pointing out that “[t]he same passage might appear in a long narrative poem and circulate independently as a lyric” (1994: 162).

<sup>17</sup> chapter 13 contains a discussion of the relationships between the six manuscript versions edited in this thesis in relation to this criterion.

the organization of a discourse colony can vary considerably; even if the order of the component parts does not influence the ‘meaning’ of a discourse colony, it can nevertheless be argued to constitute a part of its *identity*. Furthermore, also the component texts themselves exhibit internal textual variation, forcing us to account for two independent levels of variation and making it impossible to establish a single critical text without essentially creating an entirely new, editorial version of the work.

In the case of discourse colonies like recipe collections, the unstable and nebulous nature of the medieval work also has interesting and somewhat problematic implications regarding the relationship between a composite *work* represented by discourse colony texts and its individual component works. While each component text contained in a specific version of the discourse colony can be considered to belong to the colony to an equal degree, once we superimpose all of the versions of the discourse colony—each containing a slightly different selection of components—to make up a *work*, it becomes obvious that different component works are members of the colony to a varying degree, depending on how many versions they appear in. This also raises a question about the limits of the discourse colony: should we consider the colony to include all components that occur in at least one of its versions, or would it be more appropriate to include only those that occur in more than one version, viewing singular occurrences merely as accidental *hapax legomena*. In addition to comparing individual components in terms of their centrality with respect to the colony, individual versions of the colony can also be compared in terms of the combined centrality of the component works they contain. Figure 2.1 illustrates the general principle of the variable ‘centrality’ of component works in discourse colonies caused by the incomplete overlap of the different versions in terms of component texts included in them.<sup>18</sup>



**Figure 2.1:** Graphical representation of the principle of variable centrality of component works in discourse colonies.

<sup>18</sup> Figure 13.3 in chapter 13 contains a similar visual representation of the varying centrality of the recipes contained in the six versions of the *Potage Dyvers*, presented as a six-set Edwards-Venn diagram.

### 2.1.5 Material paratext or ‘bibliographic codes’

A text, from an editorial vantage, appears in its ultimate form as a linguistic or verbal event, and the act of interpreting texts consequently tends to appear as an operation we must perform on a definite and localized set of words. A more comprehensive sociohistorical view of texts, however—for example a view of texts as books, manuscripts or otherwise materialized objects—forces us to approach the issues of criticism and interpretation in a very different way, for the language in which texts speak to us is not located merely in the verbal sign systems. (McGann 1985: 190-191)

Regardless of whether we as linguists, historians or literary scholars are interested in the *text* or the *work*, we cannot escape the fact that once we approach them as readers, we will always perceive the abstract text or work through a particular material realization—the document. Since we cannot differentiate our response to a work or a text from our response to the particular material representation of that text or work, we also cannot ignore the material properties of that realization in our analysis the text or the work (Machan 1994: 66). The *work* must always remain an abstraction of an abstraction, accessible only through a semantic interpretation of a linguistic *text*, itself produced by an interpretive encounter with the physical document.

The fact that the information contained in textual documents is not limited exclusively to the textual level was already referred to in the quote from Pierazzo and Stokes (2010: 398) on page 18, and has been a frequent topic of discussion in recent years. One of the earliest and certainly one of the most vocal proponents of this view in the field of textual and literary criticism has been Jerome McGann, who in his 1991 work, *The Textual Condition*, argues that:

Textual and editorial theory has heretofore concerned itself almost exclusively with the linguistic codes. The time has come, however, when we have to take greater theoretical account of the other coding network which operates at the documentary and bibliographical level of literary works. (McGann 1991: 78)

By this “other coding network” he refers to such material and social aspects of the book as “typefaces, bindings, book prices, page format, and all those textual phenomena usually regarded as (at best) peripheral to ‘poetry’ or ‘the text as such’” (McGann 1991: 13). While McGann was writing in the context of modern literary works, the same impulse has been shared by scholars in other fields as well, as when Michael Sperberg-McQueen—one of the pioneers and leading specialists of text encoding and markup—wrote that “claiming that the only essential part of a text is its sequence of graphemes” is “a misguided and inadequate theory of texts” (1991: 35), or when Nichols (1991: 54) argued in the context of manuscript studies that the fact of the same material being rubricated differently in different manuscript versions indicates that manuscript rubrication is in fact “an artifact of

intentionality distinct from convention” and thus a carrier of information independently of the textual layer.

Since the early 1990s, this call for a more comprehensive theoretical understanding of historical documents as meaning-bearing objects has been frequently repeated by scholars representing a range of disciplines dealing with historical documents (see e.g. Markus 1997: 216, Echard and Partridge 2004b, Fraistat and Jones 2006: 106, Gants 2006: 128 and Buzzetti 2009: 48). Unfortunately, this acknowledgement of the significance of documentary features has taken a longer time to permeate actual editorial practice, even in the field of editing practical writings, a field less bound by the long tradition of critical literary editing. As Carroll (2006: 323) points out, for example Hieatt’s 1998 article on editing Middle English culinary manuscripts does not even mention the visual features of the manuscript page, and her 1996 edition of the culinary recipes found in MS Harley 5401 makes only passing reference to “underlining in red” in the recipes without providing any information on what was underlined. These omissions are symptomatic of two facts: first, most editions are designed for audiences other than linguists—in Hieatt’s case most likely culinary historians—and second, even most linguists and discourse analysts have only recently started to pay attention to the ways in which the visual and paratextual features of texts influence our understanding of their structure and pragmatic function.

The material aspects of documents and their relationship to the text were placed in a theoretical context—although very tentatively and mainly from the point of view of published literary works—by Gerard Genette in his seminal 1987 work *Seuils* (translated in 1997 as *Paratexts: Thresholds of Interpretation*), where he introduced the concept of *paratext* to describe those things in a published work that accompany the text itself and provide it with an interpretational framework, such as the cover, the title page, dedications and other prefatory material, illustrations and notes. While the applicability of many of the specific paratextual components defined by Genette to medieval manuscript texts is limited, the general concept of *peritext*, i.e. paratextual elements contained within the physical document itself, and especially the more specific concept of *material paratext*, i.e. the typography (or orthography), layout, size, choice of paper or parchment and other physical attributes of the document (Genette 1997: 7, 34),<sup>19</sup> are not only applicable but extremely relevant to the editing of medieval documents for the purposes of corpus linguistics.

In *The Textual Condition*, McGann took up Genette’s idea of material paratext, making a distinction between the “linguistic and bibliographical codes” (McGann 1991: 13) of a document: “Every literary work that descends to us operates through the deployment of a double helix of perceptual codes: the linguistic codes, on one hand, and the bibliographical codes on the other.” (McGann 1991: 77.) While both serve to convey the meanings that make up the *work*, the *textual codes* serve to communicate the structural and semantic structure of the text, and the *bibliographic codes* relate to the design and presentation of the textual content as a physical document. According to McGann (1991):

<sup>19</sup> Unfortunately, Genette does not theorize the material paratext beyond mentioning it, because as a nonlinguistic textual phenomenon, he sees it as falling outside his concerns.

This distinction, between a work's bibliographical and its linguistic codes, is fundamentally important for textual criticism, and hence for critical editing. Without making and implementing the distinction in detailed ways, textual critics cannot fully elucidate - cannot analyze - the lines of materials which descend to them. (McGann 1991: 52)

and:

By studying texts through a distinction drawn between linguistic and bibliographical codes, we gain at once a more global and a more uniform view of texts and the process of textual production. Body is not bruised to pleasure soul. (McGann 1991: 13-14)

Although McGann's distinction is widely quoted over the last two decades,<sup>20</sup> it has not gone without criticism. For example Eggert (1994) has argued that by invoking the notion of *bibliographic codes* McGann "instantly converts physical matter into textual" without appreciating the implications of the "codings' existential status as a document" (23):

What is it, after all, that documents document? They bear physical inscriptions which record prior textual processes, but while unread are only their inert residue. Their textual stasis—their status as historical record—is what calls up the editorial activity of ascertaining how and by whose hand(s) they came into existence. To take this view of document is to see that what McGann calls linguistic codes are, in the very act of being physically recorded, rendered bibliographic as well—just as the typography is bibliographic and also open to readings as a form of 'text.' McGann's distinction of the bibliographic and linguistic is initially attractive but finally, for me at least, blurs the more fundamental document-text distinction. (Eggert 1994: 23)

Eggert's criticism is very much in line with the multi-leveled concept of textual objects adopted in this thesis, highlighting the fact that the linguistic *text* and the *material paratext* (or bibliographic text, to use McGann's terminology) are not parallel but sequential phenomena. As Nichols (1991) has formulated it using concepts defined by Emmanuèle Baumgartner, medieval (as well as modern) literary *works* can be seen to simultaneously inhabit both the "textual space" and the "manuscript space". From this viewpoint, the contribution of both Genette and McGann lies in their observation that the physical level of the *document* contains more information than what is traditionally considered to belong to the linguistic level of the *text*. Or in other words, not all of the linguistically significant information contained in the *document* can be represented as *text*.

In the context of this thesis, the material representations of this 'surplus' meaning that is encoded in the document but defies linguistic interpretation and representation as *text*, is referred to by the term *material paratext*.<sup>21</sup> The term *visual paratext* is used to refer to a subset of material paratext that is encoded as visual

<sup>20</sup> See for example Moffat 1998: 44, Kirschenbaum 2002: 43, Flanders 2006: 142, Fraistat and Jones 2006: 106, Robinson 2006: 82, Dahlström 2009: 33 and Sutherland 2009: 20.

<sup>21</sup> The parent concept of *paratext* is not very useful in the context of the present edition, as the absence of an author as the defining principle of the *text*, the linguistic paratext is placed onto an equal



markings made on the document but is not interpretable in linguistic terms.<sup>22</sup> Features of medieval manuscripts that are a part of the material paratext but not of the visual paratext include codicological features of the material, size and format (roll, codex, sheet, etc.) of the document, as well as its binding. As Stokes (2010) points out, even these features—often overlooked by editors—are important indicators of the function and status of the document: “a pocket-gospel was probably made for personal use, whereas a large-format bible may be an assertion of wealth and power” (237).<sup>23</sup>

In terms of *visual paratext*, the features employed in medieval manuscripts are different but no less varied than those mentioned by Genette for modern printed texts. As Nichols (1991) has observed, medieval manuscripts may not have had “the variety or the convention of typographical variation that we are accustomed to” but instead used two forms of graphical highlighting not commonly used in modern printed texts, namely decorated initials and rubrication (53).<sup>24</sup> Machan (1994) also sees medieval texts as utilising their material context in more diverse ways than modern ones, using “highly expressive features of layout and design that the manuscript producers could consciously manipulate” for pragmatic purposes (65).

Specific visual paratextual features that have been argued by different scholars to serve a communicative function or “carry a cultural significance” (Taylor 2004: 78) in medieval manuscripts include:

- layout (including lineation and the use of marginalia) (Machan 1994: 65, Taylor 2004: 79, and Pierazzo and Stokes 2010: 398);
- illuminations and illustrations (including their presence or absence) (Machan 1994: 65, and Taylor 2004: 79)
- enlarged and decorated initials (Machan 1994: 65, Markus 1997: 216, Robinson 2006: 82, and Nichols 1991: 54)

---

standing with what Genette would consider the ‘text itself’. For this reason, paratextual elements like incipits, explicits, headings and tables of contents, are here considered functionally specialized components of the *text*, instead of being merely “written and visual signs marking the boundaries of the textual space” (Nichols 1991: 57).

<sup>22</sup> It should be noted that the concept of material (and visual) paratext is not here used to refer to the actual, physical ink or pigment on the physical document page (which are features of the *document*), but rather to an abstraction of those markings, ontologically parallel but separate from the *text*: just as *text* is an abstract typologization of strokes of ink into abstract letterforms, the *material paratext* is an abstract typologization of actual physical features of the *document* into abstract entities.

<sup>23</sup> These kinds of features are also the most difficult ones to represent in an edition, even a digital one. Stokes (2010: 237) suggests images and video—such as the resources available through the Evellum project (<<http://www.evellum.com/>>)—as a means to mitigate the problem, but this naturally requires significant co-operation from the holding institution of the document. Because of the unavailability of publishable images of the documents edited in the current edition, this information is provided in the textual descriptions of the manuscripts provided both in the headers of the digital documents themselves (see appendix A) and as a part of the thesis (see section 9.2).

<sup>24</sup> Although it should be pointed out that for example the practice of using a different typeface to highlight foreign language words or headings can be seen to go back as far as the Anglo-Saxon customs of using Caroline minuscule for Latin and Insular minuscule for Old English, and of using majuscule scripts of different sizes for indicating textual hierarchy (Denholm-Young 1954: 24; Hector 1966: 51; Parkes 2008: 57). Considering that in the sixteenth century it was not uncommon to use italic script to highlight signatures, quoted matter (whether in Latin or not), headings and marginalia in documents otherwise written in secretary (Hector 1966: 61-3), it is likely that the typographical conventions of early printed books were in fact inherited from manuscripts.

- rubrication and other use of colour (Markus 1997: 216, Nichols 1991: 54, 61, and Taylor 2004: 79)
- decorative flourishes (Machan 1994: 65)
- different scripts or letterforms (Markus 1997: 216, Taylor 2004: 79, and Pierazzo and Stokes 2010: 398); and
- hand size (Markus 1997: 216).

All these features have been seen to either “accent or counterpoint a text in any number of ways and thereby contribute to the reader’s perception—even construction—of the *res* behind the *verba*” (Machan 1994: 65), forming a part of the communicative apparatus of the document, or to provide us—students of history, language and literature—with important insights into the pragmatic and cultural functions of the text (Pierazzo and Stokes 2010: 398). For example the presence or absence of ornamentation, rubrication and enlarged initials allows us to differentiate between “economical and uneconomical manuscript formats” and are “indicative of respect or potential use” (Toon 1991: 88), while “[g]losses, signs of wear, and other codicological details can reveal a great deal about how and how extensively a manuscript was used” (Diehl 2004: 61), and marginal annotation, even when it has nothing to do with the text itself, can reveal things about the reception of the text and the ways in which it was used.<sup>25</sup>

The importance of taking the material paratext of Middle English manuscripts into account—and somehow representing it in editions—is emphasized by the fact that we still have a very tenuous grasp of the significance of these various features. Since we do not have “definitive reference works explaining the pragmatics of medieval bibliographic codes” (Machan 1994: 184-5), the only way of uncovering the pragmatic functions and cultural implications of the visual paratext of the medieval manuscript is to record and analyse them systematically:

If there is one area in Middle English textual criticism that needs particular work, [...] it is the determination of the meaning and relevance of medieval bibliographic codes. And if there is one area in Middle English editing in need of particular work, in turn, it is the representation of these codes.  
(Machan 1994: 186)

Unfortunately, the predominant tradition of critical editing (described in section 3.1) and print publication of editions has established the practice of converting the native paratext of the medieval manuscript to the paratextual format of the modern printed book as a normal practice, thus severely distorting the textual reality of the original (Cerquiglini 1999: 27). According to Machan (1994), the problem lies in the fact that whatever the critical editor’s feelings towards the material features of Middle English texts, he has been constrained by the received tools of humanist textual criticism, all of which are essentially *lexical* in nature and thus inapplicable for the representation of the material document.

However, the developing possibilities of digital publication media, together with the increasing theoretical awareness of “the importance of the materiality of the book in generating meaning”, displayed even by textual critics like Jerome

<sup>25</sup> As an example of this, Toon (1991: 90) mentions marginal doodles made in dry-point to a collection of Latin sermons, which he interprets as an indication of a bored audience, and as an indication “that the manuscript was laid out on a table with pupils, scribes, or colleagues gathered round”.

McGann, D. F. McKenzie, John Barnard, and James McLaverty (Greetham 1992: xix, Eggert 1994: 14), have over the last two decades gradually made it possible to develop scholarly editions in which all of these features are recoverable—at least to some degree—by the user and can be analysed in relation to the textual content of the document. Since this thesis shares the view of Robinson (2009: 47) that the appearance of the manuscripts—i.e. their visual paratext—is inherently significant and “part of their utterance”, and therefore worth annotating to a relatively great detail, the current edition aims to formally encode not only the text, but also the visual paratext of the edited documents.

## 2.2 Context

[T]exts are produced and reproduced under specific social and institutional conditions, and hence [...] every text, including those that may appear to be purely private, is a social text. This view entails a corollary understanding, that a ‘text’ is not a ‘material thing’ but a material event or set of events, a point in time (or a moment in space) where certain communicative interchanges are being practiced.

(McGann 1991: 21)

As McGann and several other textual scholars have pointed out, textual objects are always both created and received in a specific *context*, which influences practically all aspects of their realization first as a linguistic *text* and finally as a *document*. Being thus encoded into the document as various kinds of presuppositions, the context also forms the interpretive framework in terms of which the document is intended to be understood. In spoken communication or roughly contemporary written communication these presuppositions are shared to a large extent by all participants of the communicative act and the contextual nature of communication can often be ignored, but in the case of historical texts, it becomes acutely important for understanding the meaning of the text—i.e. reconstructing the *work* on the basis of the *text*—and even for decoding the text from the physical document.

### 2.2.1 Layers of context

Although *context* is often treated as a single concept, its role in the communicative situation becomes easier to understand if we break it down into several types or layers. Leckie-Tarry (1993, 1995) has proposed a useful layered model of context—similar to the three-tiered approach to the study of discourse suggested by Fairclough (1992: 4)—based on the *register theory* (see subsection 2.2.2 below) developed by M. A. K. Halliday and other scholars,<sup>26</sup> which posits three levels of context involved in the meaning-making process: the *context of text*, the *context of situation* and the *context of culture* (Leckie-Tarry 1995: 17, 159).

<sup>26</sup> See for example Halliday 1978 and 2004, Halliday and Hasan 1976, Martin 1992, Mathiessen 1993, Eggins 1994, and Eggins and Martin 1997.

In this classification, the *textual context* refers to the *discourse* surrounding the textual object—both in the wider sense of all the other textual activity going on in the society surrounding the text, and in the narrower sense of other texts physically and logically connected to the text. Discourse, from this point of view thus becomes a superordinate concept for *text*, which is conceived of as the product or realization of the discursive process (Fairclough 1992: 3). The *situational context*—a concept originally introduced by the anthropologist Bronislaw Malinowski (1923: 306–309) and subsequently elaborated by other scholars—refers to the immediate environment of the communicative event encoded by the text, both in terms of the physical situation and the topic of communication, the interpersonal relationships of the discourse participants, and the mode of communication (Leckie-Tarry 1995: 159).<sup>27</sup> Finally, the *cultural context*—also originally introduced by Malinowski (1935: 18)—refers to the wider sociocultural structures, ideologies and ‘thought-styles’<sup>28</sup> within which the communicative situation is taking place. To these established ‘layers’ of context, could be added the wider *linguistic context* in which the textual object is produced, referring to the language system forming the *langue* conditioning the *parole* of the *textual context*—including the text being contextualized, as well as the *material context* that defines the ways in which the *text* gets encoded into a physical document and the features of *material paratext* that are available for encoding meanings.

In the case of the culinary recipe collections edited in this thesis, their cultural context can be seen to be constituted by the medieval ‘thought-styles’ concerning food and nutrition, including its social and religious roles. The more concrete situational context, on the other hand, is centred around the everyday life of the medieval aristocratic or genteel household, with a focus on the organization of the medieval kitchen, on the structure and logic of the medieval meal and on the institution of the medieval feast. The immediate textual context of these recipe collections in the broad sense is made up not only of the other texts physically bound in the same documents as the recipes, but also the whole body of Middle English recipe collections surrounding them, while their wider linguistic context is made up of the general language and literacy practices of fifteenth-century England. Finally, their material context is made up of the technology and practice of writing as a physical activity and of book production in general. In order to con-

<sup>27</sup> The division of the situational context into these three components was first made by Halliday, McIntosh and Strevens (1968: 78–94). Halliday (1978: 33) concisely defines these three components as 1) the *field* of discourse, or the institutional setting in which the piece of language occurs, including not only the subject matter but also what the participants are engaged in doing; 2) the *tenor* of discourse, or the relationships between the participants, covering not only their degree of formality but also the degree of permanence and the emotional charge invested in the relationship; and 3) the *mode* of discourse, or the channel of communication adopted and the role of language in the interaction.

<sup>28</sup> The concept of ‘thought-styles’, originally defined by Crombie (1994), is here borrowed from the Scientific Thought-styles project of the Research Unit for Variation and Change in English at the University of Helsinki, where it is used in the context of medieval and early modern science to refer to “the underlying scientific concepts, objects of enquiry, methods, evaluations and intellectual commitments related to the epistemology of science” (Pahta and Taavitsainen 2011; see also Taavitsainen and Pahta 1995: 519, and Pahta and Taavitsainen 2004). In the context of medieval culinary recipes, it could be seen to encompass the medical and dietary theories current in late medieval Europe, the religious doctrines, views and restrictions concerning food, and the social aspect of food as a constituent and expression of the medieval social system.

textualize the recipe collections edited in this thesis, part II of this thesis provides information on the different levels of context mentioned above, as well as the general linguistic and literate environment in which they were produced and used.

### 2.2.2 Register and genre

The relationship between the cultural and situational contexts and their linguistic realizations is described by a variety of linguistic theory known as *register and genre theory*:

What unites the work of linguists working on register is the centrality of text viewed in its context of social situation. Register *entails* text and implies a relationship between text and context. Register analysts explore the link between linguistic expression and social situation, with a view toward explanation. (Biber and Finegan 1994: 7)

The basic premise behind register theory is that a text always carries with it—as a part of itself—aspects of the context in which it was produced (Eggins 1994: 7), the linguistic system’s adaptation to the diversity of contexts being seen as *register variation* (Mathiessen 1993: 235-6). The concept of *register* is thus simply a theoretical explanation for the common sense observation that we use language differently in different situations (Eggins and Martin 1997: 234), or as Beaugrande (1993) puts it, “a set of beliefs, attitudes or expectations about what is or is not likely to seem appropriate and be selected in certain kinds of contexts” (18). It is the aim of register theory, therefore, to uncover the general principles governing the variation of language according to the type of situation, and to gain an understanding of which situational factors determine which linguistic features (Halliday 1978: 32).

It should be noted that the notion of register can be seen both as a form of prediction—given information about the social context of language use, we can predict a great deal about the language that is likely to occur (Halliday 1978: 32)—and of “contextual *deduction*” (Eggins and Martin 1997: 236), enabling the elucidation of the context based on the resultant text. It is precisely this latter application of register theory that makes it a useful tool for studying medieval recipes and other historical texts that occur in a context that is likely to be highly specific yet sparsely documented. Thus a *register* is here seen as the response of the semantic system to a certain configuration of situational variables; a model of the semantic responses that a given situation is likely to elicit from a discourse participant. Although the register of a text manifests itself as a particular selection of words and structures, it is defined in terms of meanings; a register “is not an aggregate of conventional forms of expression superposed on some underlying content by ‘social factors’ of one kind or another”, but rather it is the selection of the *meanings* expressed “that constitutes the variety to which a text belongs” (Halliday 1978: 111).

The concept of *text type*, which has been used in widely varying ways by different scholars is here seen simply as this same phenomenon viewed from the direction of language, referring to the discursive features<sup>29</sup> shared by those texts

<sup>29</sup> The concept of *discursive feature* is here used to refer not only to linguistic (morphosyntactic and lexical) features of texts, but also to their textual and paratextual features (information structure,

that a given situation is likely to prompt. However, it should be noted that since registers (as well as *genres*, discussed below) are defined by language-external features and text types by language-internal features, there is no *a priori* coupling between them. This in turn means, as Biber (1988: 6) has pointed out in connection with genres, that linguistically distinct texts belonging to the same register and genre can represent different text types and linguistically similar texts in different registers and genres can represent the same text type.

While *register* is usually used to describe the influence of the *situational context* on the text, the concept of *genre*—which already has a long history in literary studies—has been used in a similar way to refer to the realization of the *cultural context* in the structure and textual features of the textual object. Although Halliday (1978) already acknowledged the existence of a separate generic structure derived “from a higher-level semiotic structure” (134) outside of language, it was Eggins and Martin (Martin 1992, Eggins 1994 and Eggins and Martin 1997) who properly integrated the concept of genre into the framework of register theory. They also noted that the commonly used linguistic definition of genre differs from those used in literary studies in two respects: first of all, linguistic genres include all kinds of language use, and second, they are defined functionally in terms of their social purpose. In their own words, “different genres are different ways of using language to achieve different culturally established tasks, and texts of different genres are texts which are achieving different purposes in the culture” (Eggins and Martin 1997: 236).<sup>30</sup> Although the basic premise of this definition—formulated by Eggins and Martin (1997: 236) as a suggestion that “texts which are doing different jobs in the culture will unfold in different ways, working through different stages or steps”—is now generally accepted within the fields of pragmatics and discourse analysis, the concept of genre has still been difficult to integrate with the older concept of register (Swales 1990: 41).

Following Eggins (1994: 9–29), I will here define *genre* as the overall function of the text as a form of *recognized cultural activity* (such as negotiation or instruction). According to Leckie-Tarry (1995: 145), this kind of “macro-proposition” that defines “what the discourse is about” is in many texts necessary for creating meaning. Such texts are thus accessible only through reference to the context of culture. Genres not only codify the registers that are considered suitable for achieving a given goal<sup>31</sup>, but also give a blueprint for the structured process of individual linguistic acts for achieving that goal Eggins (1994: 36). For every such activity, social convention (the *context of culture*) establishes a set series of steps that we go through in order to achieve the purpose of that activity. This series of steps is known by linguistic genre theorists as the *schematic structure* of the genre (Eggins 1994: 36). Thus *genre* can be seen to operate similarly to *register*,

textual organization, visual layout, etc.).

<sup>30</sup> Also Swales (1990: 33–45) presents a very similar functionalist definition of genre—based on a combination of genre definitions used in various fields, such as folklore studies, literary studies, linguistics and rhetoric—for use in the field of discourse studies. According to him, a genre is a “class of communicative events” in which language plays a significant part. The principal criterial feature that turns a collection of communicative events into a genre is “some shared set of communicative purposes”.

<sup>31</sup> This relationship is also very similar to the relationship between Bakhtin’s (1986) concepts of *primary* and *secondary genres*, which correspond quite accurately to the concepts of *register* and *genre* used here.

but in relation to the context of culture instead of situation, enabling us to “describe [...] the interface between the socio-cultural world and textual form” and the “ways in which texts and the social agents which produce them construct and are constructed by the social and the cultural” (Kress and Threadgold 1988: 216).

It is important to note, that the text produced in a given situation within a given culture is always a combination of or compromise between the demands and possibilities of the situation and those of culture (Mathiessen 1993): representatives of two different cultures are not likely to produce a similar response in a similar situation due to differences in the generic repertoires customary to them. It must also be noted that genre and register do not *dictate* the text but merely present a range of functionally and conventionally relevant options for the speaker or writer, i.e. the *text types* appropriate for the specific genre and register. The individual text is thus not a direct product of the register and generic structure, but also allows for modulation by the individual *style* of the speaker or writer. The resulting *text*, finally realized through the various features of the lexicogrammatical system as *language*, is thus a product of all the various contextual factors and requires reference to all of them in order to be understood.<sup>32</sup>

### 2.2.3 Context of historical texts

In the case of historical written texts, the context is a problematic issue from a communicative point of view, as all of the three levels of its original context are available to the modern reader only in a very incomplete and indirect form. An instance of spoken communication can usually assume all of the three levels of context to be more or less shared by the communicating parties, and even contemporary written communication can usually rely on at least the textual and cultural contexts of the writing and reading situations to be shared to some degree. In contrast, historical texts are not only read in a very different cultural and situational context than the one they were written in, but may also originate in situational contexts the parameters—or even the very existence—of which we have no idea of, unless they happen to have been documented by surviving documentary or archaeological evidence. Even the textual context of historical writing—our most important source for all the levels of context—is incomplete at best, and practically non-existent at worst. This means, as McGann (1991) puts it, that “all we can do is make imaginative attempts at reconstituting or approximating it for later persons living under other skies” (83).

This problem becomes especially acute in the case of nonfictional—and especially practical—writing, as they are not only produced within a context, but unlike fictional texts, they by definition frequently rely on this context and the *shared information* implicit in it for their meaning (Werlich 1983: 17–21, 42–3). The best we can thus achieve in the case of practical historical writing is a limited and always somewhat uncertain interpretation of the text encoded in the document, and an even more uncertain interpretation of what his text means, i.e. the work represented by the text. But even this requires that we first of all do our best to rec-

<sup>32</sup> It should be noted that the definition of *text* used here differs from that of Halliday (1978: 135) in that although he sees the context to be finally realized on the lexicogrammatical level, he uses *text* to refer not to a string of clauses and sentences, but rather to a *semantic* concept, making it roughly equivalent to the *work* in the scheme described here.

ognize the particular differences between the production and reception contexts of the document, and second, keep in mind that even those aspects of the original context that we think we recognize, may in fact have undergone “silent but spectacular changes” (Modiano, Searle and Shillingsburg 2004: xi) that significantly skew our perception of the meanings and pragmatic functions of the documents and their texts. In practice, this means that not only do we have to familiarize ourselves with what little is known about the context of the texts we are reading or editing on the basis of previous research, but we must also keep an open mind and be prepared to question this received wisdom, which is after all based mainly on evidence provided by the very documents that we are editing.

#### 2.2.4 The contextualized textual object

Although most of register theory has focused only on the context of situation at the expense of the other contexts, this thesis follows Leckie-Tarry (1995: 17, 34) in seeing the selection of semantic and linguistic structures in the textual object as dependent on the interaction of all the different levels of context. The levels of context can be seen to relate to the different levels of the textual object along two dimensions, forming a structure which can be viewed as a stratified ‘system of systems’. As Mathiessen (1993) has observed, each of these systems has its own internal organization, but is also related to the other systems by being situated in a ‘realizational chain’, where each system realizes a higher system and is in turn realized by a lower-order system: “This chain of inter-stratal realizations bridges the gap between the semiotic in high-level cultural meanings and the material, [...] through a series of intermediate strata” (226). For the purposes of this thesis, I will propose a model of contextualized language that combines the slightly disparate approaches of Halliday, Martin and Leckie-Tarry by positing a two-dimensional chain of realization, organized along a ‘cline of instantiation’ in one dimension and a ‘cline of realization’ in the other.<sup>33</sup>

These layers of context and the discursive practices represented by *genre* and *register* thus interact with each other and influence the production and reception of the textual object on every level, as *Figure 2.2* tries to show in schematic form. The cultural context not only influences the range and parameters of the situational contexts in which textual activity can take place, and thus the kinds of ideas or *works* that are conceivable within that culture, but also defines the overall functions for which writing is being used in the form of *genres*, which codify the established textual means of fulfilling these functions, affecting the realization of the work on the lexicogrammatical level in concert with the situationally defined *register*. The *situational context*, in turn, defines the appropriate linguistic means of realising the culturally defined purpose of the *genre*, i.e. the appropriate *register(s)*. The *genre* and *register*, as collections of communicative conventions, or *discursive practices* (Fairclough 1992: 5), thus together determine the appropriate ways of encoding the envisioned work on the abstract linguistic level of the *version*, while the abstract *linguistic context*, (*‘langue’*), via its realization as the surrounding textual context (*‘parole’*) with its dialectal and orthographic norms, guide the realization of this *version* as a specific *text* with the appropriate orthographic realizations of

<sup>33</sup> A similar approach has also been taken by Mathiessen (1993), whose thinking seems to proceed along much the same lines as the one presented here, as different as the resulting models may be.



the semantically appropriate grammatical and lexical items. This text is then encoded as a document according to the culturally and technologically determined conventions of handwriting and book production, or the *material context*, which also defines the appropriate *material paratext* for the culturally and situationally determined textual context, resulting in the multiply coded *document*.

In this model the concept of *text type*—which is not a part of the matrix presented in Figure 2.2, is a categorical concept that can describe groups of textual features on several levels. A *generic text type* would refer to the *discursive features* shared by the texts produced as instantiations of a certain *genre*, while a *situational text type* would refer to the discursive features shared by texts produced under the same broad situational conditions.<sup>34</sup> The intersection of these two levels of text type would then form a more specific level of text types ('registerial text types'), each describing the discursive features shared by texts performing a certain culturally defined communicative task (i.e. belonging to a certain genre) in a specific situational context. This kind of hierarchical model of the interplay between context and text allows textual features to be examined on different levels, the most appropriate level being determined by the scope of the individual research question.

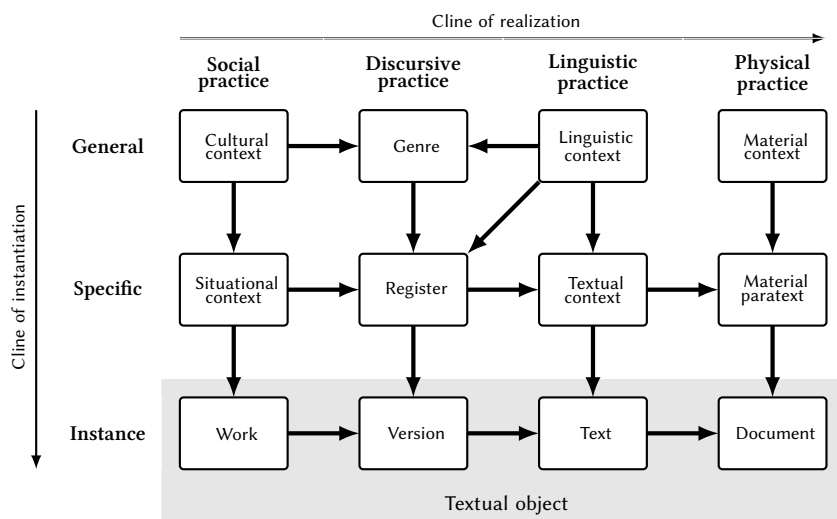


Figure 2.2: The relationships between the different levels of context and their influences on the different layers of the textual object.<sup>35</sup>

The omission of an *author* from the preceding description of the interplay be-

<sup>34</sup> One could define a further sublevel of text types, pertaining to the individual components of the situational context (as defined by Halliday (1978: 33)), referring to the discursive features shared by texts relating to the same *field*, *tenor* or *mode*. Similarly, it would be possible to define a superlevel of *cultural text type*, which would describe the discursive features shared by all texts produced in the same cultural context.

<sup>35</sup> Although the relationships of realization are represented as single-headed arrows, we should not forget that the relationships between the categories in fact work both ways, and changes in textual patterns can be interpreted as both reflections of changes in society and as construing social change (Mathiessen 1993: 251).

tween text and context is entirely intentional. Partly this is because of the undesirable semantic baggage attached to the term by literary textual scholarship—which could be avoided by merely using a more general term like *writer*—and partly because the influence of his or her personal idiosyncracies on the produced textual object can be seen simply as another aspect of the context in which the textual object was produced. In this sense the approach taken in this thesis is similar to that of Machan (1992), who sees—in opposition to the traditions of literary textual criticism—*textual authority* as emerging not from the creative genius of an *author*, but rather “from all the collective aspects of a work’s production, transmission and reception: its sources and compilation, the number and chronological spread of its copies, the physical and lexical similarity or dissimilarity of these copies, and the intended social function (when recoverable) and actual social function of these copies” (13). While the concept of author is not included in this conceptual scheme, it will play a role in the more practical discussion of the transmission of manuscript texts below.

## 2.3 Authors, scribes and textual transmission

Amont incomparable, l’auteur, grand par définition, tranche absolument, par l’unicité de sa conception, l’opacité de son œuvre (argument de la *lectio difficilior*), la qualité de sa langue, avec la diversité scribale, ignorante et sans dessein, qui pluralise l’œuvre, en banalise l’expression, appauvrit la langue.<sup>36</sup> (Cerquiglini 1989: 90-91)

As Machan (1994: 54-6, 131-2) has pointed out, the humanist conception of *authorship*—a historically important concept because it underlies much of modern critical editing—satirized by Cerquiglini in the above quote is largely inapplicable to Middle English writing, as it is intimately tied to the appropriation and commodification of discourse in the renaissance period, implying an ownership and responsibility for a text (Foucault 1979: 148). In contrast, according to Machan, medieval people saw the content of the work, its *res* as its significant aspect, not the *verba*, or the individual rendering of that content by an individual author. Neither was the precise identity of the author of a book they were reading or quoting very important, as is indicated by “hundreds of unattributed medieval compositions” (3). Cerquiglini (1999) goes even further and argues that “although the emergence of the figure and practice of the writer can be shown starting in the fourteenth century”, the ‘author’ is not a medieval concept and the expression *medieval author* seems to him “a functional anachronism” (8):

<sup>36</sup> “The author, great by definition, and unique, the most pre- of pre-production by the unity of his conception, the opacity of his work (the argument of *lectio difficilior*), and the quality of his language stood in sharp contrast to scribal diversity, ignorant and purposeless, which pluralized the work, trivialized its expression, and impoverished its language.” (Translation in Cerquiglini 1999: 61.)

As with the printing press, the Renaissance set new ideas in place, securely albeit in a halflight. The emergence of the author in the period from the sixteenth to the nineteenth century is a complex but well-known phenomenon that falls into the realm of what might be called 'internal literary history'.<sup>37</sup> (Cerquiglini 1999: 8-9)

The *medieval* concept of authorship—or an *auctor*—was quite different from the modern concept of a literary *author* used in discussions of editorial theory. The medieval *auctor* was not only a writer, but also an 'authority', "one who writes works of 'truth', which were worthy of imitation" (Caie 2008: 20), "someone not merely to be read but also to be respected and believed" (Minnis 1984: 10):

The *auctor* may profitably be regarded as an accolade bestowed upon a popular writer by those later scholars and writers who used extracts from his works as sententious statements or *auctoritates*, gave lectures on his works in the form of textual commentaries, or employed them as literary models. Two criteria for the award of this accolade were tacitly applied: 'intrinsic worth' and 'authenticity'. (Minnis 1984: 10)

Of these criteria, 'intrinsic worth' referred to the conformance of the work, in some way, to Christian truth—the Bible being "the authoritative text *par excellence*"—and 'authenticity' to the text being "a genuine production of a named *auctor*" (Minnis 1984: 11). In an age which saw itself as inherently inferior to the glorious past of antiquity, this meant that no contemporary writer—much less a vernacular one—"could decently be called an *auctor*" (12).

Although writers of Middle English texts may have "lacked the status, dignity, and authority to be considered *auctores* in the medieval sense, that does not mean that they were not authors in any sense" (Jacobs 1998: 6). Jacobs (1998) has argued for a conception of authorship that does not depend on any artistic originality of material or even of its treatment, but rather "on an intention to communicate on a wider scale than that of immediate personal interaction" (7) and on consciously "generating a composition" (6). This should not be taken to mean—as some approaches that try to de-emphasize the author sometimes assume—that a medieval text producer was an average or typical member of his or her society with no special qualities. Considering the non-trivial status of literacy skills in the medieval context (see section 6.4), this would most likely be as mistaken as the concept of an author as a genius. The actual person who conceived—in whatever sense—a text most likely had several properties that him (or her) apart from his (or her) contemporaries: his craft and technique as a writer, the bilingual skill of a translator, his expertise in the language peculiar to his topic of choice, and his possible knowledge and skill in the technical subject itself (Moffat and McCarren 1998: 49–50).

It is also important to keep in mind that the figure of the *author* is very much connected to the idea of the *origin* of a text, which in itself is not a very useful

<sup>37</sup> As Cerquiglini points out, this later emergence of the modern author as an artistic genius also has to do with several developments in "external literary history", like the waning of royal and patronage and private sponsorship, the demands—beginning in the 1720s—for financial autonomy, mainly from the humbler 'rank-and-file' of writers who wanted to get some profit from the sale of their books, and conflicts in the relations between authors and booksellers (1999: 9).

concept in the context of medieval text production. Vernacular medieval writers, of whom Caie (2008) uses Gower as an example, saw themselves primarily as compilers of existing material, not as authors of original content. In the Middle Ages, the Senecan image of the bee was often invoked to describe the work of the compiler: “the bee gathers nectar, arranges it into cells and creates honey, thereby borrowing, rearranging and coming up with something new” (20). While this difference between the modern and the medieval author has not always been appreciated by editors and textual scholars, for example Cerquiglini (1999) has argued that the medieval conception of the ‘work’ and especially of its authorship was significantly different from ours in this regard: the fact that someone had to be the first to write down a given text “was probably less important” than the “continual rewriting of a work that belonged to whoever prepared it and gave it form once again” (33).

### 2.3.1 Conceptualising the author

In deconstructing the text-critical concept of *author* as a source of *textual authority*,<sup>38</sup> Foucault (1979) has observed that this concept is in fact distinct from the *writer* of the text, not referring to any real individual but to a *function* “characteristic of the mode of existence, circulation, and functioning of certain discourses within society” (148). He also points out, that while all texts have a *writer*, it is only certain texts “that are endowed with the ‘author function,’ while others are deprived of it”<sup>39</sup> (1979: 147-8), highlighting the fact that it does not “develop spontaneously as the attribution of a discourse to an individual” but is, “rather, the result of a complex operation which constructs a rational being that we call ‘author’” (150). Nor is this author-function historically stable: the types of texts that have culturally required an attribution to an author have varied in the course of history. For example texts that we would consider ‘literary’, such as narratives, stories, epics, tragedies and comedies “were accepted, put into circulation, and valorized without any question about the identity of their author”, and texts that are nowadays considered to derive their authority entirely from their content and be evaluated anonymously, like scientific writing and other explanations of the world, were in the Middle Ages accepted primarily by virtue of being associated with a known authority (149).

Thus, unlike implicitly assumed by textual critics (see subsection 3.1.1), “the author is not an indefinite source of significations which fill a work; the author does not precede the works, he is a certain functional principle by which, in our culture, one limits, excludes, and chooses” (Foucault 1979: 159), a projection of “the operations that we force texts to undergo, the connections that we make, the traits that we establish as pertinent, the continuities that we recognize, or the exclusions that we practice” (150). Or as McGillivray (1994) formulates from a more editorially oriented point of view, the author as a source of textual authority is “a way of limiting the potentially inexhaustible combinations of readings, and hence of meanings, of the various manuscripts, of reducing the babble of competing voices

<sup>38</sup> The significance of authority and the concept of the *author* in traditional critical editing will be discussed in more detail in subsection 3.1.1

<sup>39</sup> As examples of discourse types that obviously have a writer but do not have an ‘author’, Foucault mentions letters, contracts and writings on the wall.

to a single voice” (177). In the context of anonymous vernacular practical writings like recipes, this traditional ‘authorising’ function of the author is thus clearly not a valid editorial tool, as we will see in chapter 3.

In order to account for the fact that the production of every text—even those not endowed with the Foucauldian *author function*—nevertheless involves some kind of human agency, Fairclough (1992: 78) has advocated an approach that deconstructs the compound concept of a *text producer* into several positions. As one alternative he proposes a threefold division of the authorial role—originally suggested by Goffman (1981: 144) in the context of spoken language—into:

- 1) the *animator*, or the actual writer of the physical text,
- 2) the *author*, or the person who puts the words together into a text, and
- 3) the *principal*, whose position is represented by the words.

These roles quite conveniently correspond to our conceptualization of the textual object into *document*, *text*, *version* and *work*, each of the authorial positions being responsible for a different step in the movement of the textual object along the realizational chain: the *principal* being responsible for the abstract and ideational *work*, the *author* for its realization as a lexicogrammatical *version*, which the animator then formulates orthographically as a *text* while fixing it onto a *document*. While all of these positions could be fulfilled by a single person,<sup>40</sup> it seems that there was a carefully maintained distinction between the activities of *dictare* (the composition of a work) and *scribere* (the writing of a text) (Machan 1994: 141, 175). Although the concept of *author* is principally associated with textual *production* (see subsection 2.3.2), the nature of manuscript documents means that the position of *animator*, and in many cases also the *author*, are invoked also when a text is copied.

### 2.3.2 Production of manuscript texts

Unlike in the modern period, the composition of texts was in the Middle Ages seen as an oral process, separate from their recording in writing on a physical medium, the former being referred to by the term *dictare*, while the term *scribere* generally referred only to the physical act of putting pen to parchment (Fleischman 1986: 20). As Machan (1994) points out, the distinction between these two activities was carefully maintained—“both socially and institutionally” (175)—throughout the Middle English period, and they were often carried out by different individuals, people like Hoccleve, who performed both activities, being an exception rather than the rule (141). In addition to these two activities, of which the first can be seen to encompass both the roles of the *principal* and the *author*, Machan (1994: 143) has also postulated two other intermediate activities that reflect the accumulative and ‘recycling’ nature of medieval textual production: *commentary* and *compilation*.

The roles performing these two additional activities can be seen as specialized forms of the *author* role suggested by Fairclough: both of them involve the rendering of a received ideational *work* into a linguistic *version*, but with the added function of either expanding on the original content (*commentator*) or reorganis-

<sup>40</sup> Machan (1994: 141) mentions Hoccleve as an example of an individual who habitually performed all of these roles himself.

ing and recontextualising it in some way (*compiler*).<sup>41</sup> The different manuscript versions of the *Potage Dyvers* can in fact be seen to exemplify both of these additional roles. The activity of a *compiler* is evident in the various reorganizations and selections made in the copying of the different versions the recipe collection, where he or she has either added material from another source, omitted some of the recipes in the exemplar or reordered them in some way. The activity of a *commentator*, on the other hand, can be seen in the changes made to the content of the recipes as they are copied, whether based on personal taste or on the predicted availability of ingredients.

The encoding of meaning in the textual object did not always end with the scribe depositing it into a document. The study of ‘formal’ manuscript annotation, like rubrication and illumination—the *visual paratext* of manuscripts—has revealed that the production of medieval manuscripts was a multi-phase process involving several craftsmen, each depositing their own layer of meaning onto the document (Nichols 1991: 48). Nichols (1991: 47) has introduced the concept of *manuscript matrix* to refer to the context of this mode of textual production, within which we need to analyse medieval literature, and by extension, also other kinds of textual objects: “Literature was not simply discourse, it also implicated the body and a whole sociological infrastructure of production: the scribe in the scriptorium, the rubricator, and artists specialised in decorated initials and miniatures.” (Nichols 1991: 47).

### 2.3.3 Scribal copying

The above mentioned features of manuscript transmission have editorially significant implications for the relationships between the different levels of the textual object that set vernacular medieval texts apart from Biblical and classical Latin texts on the one hand and modern printed works on the other, on which modern textual theory is largely based. As Eggert (1994) reminds editors, medievalists have for long known “that the usual form of publication in the Middle Ages, hand-copying [...], necessarily meant that the normal rather than aberrant form of textual existence was one of unending variation” (72). Traditionally, however, this variation between different versions of a work has been seen as an indication of *corruption*, resulting from the failure of scribes to accurately copy their exemplars.<sup>42</sup> As McIntosh, Samuels and Benskin (LALME) point out, these assumptions reflect the early medieval convention of *litteratim* copying, associated with Latin texts that provide “very little scope for departure from the precise form of the original” because of the highly inflected nature of the language and the paucity of alternative spellings (1:29), and are not necessary applicable to late medieval vernacular texts, which were copied according to very different scribal conventions.

With the resurgence of English as a written language in the mid-14<sup>th</sup> century, there seems to have been a change in copying habits towards a less verbatim practice, for which McIntosh, Samuels and Benskin (LALME: 1:29) see three contribut-

<sup>41</sup> In cases where the commentator or compiler supply significant new ideas not contained in the received material, they could also be conceived of as a partial *principal*.

<sup>42</sup> The entire methodology of Lachmannian stemmatics (see section 3.1) is based on the idea of scribes making unique mistakes in their copying, which then get copied in subsequent generations of manuscript copies.

ing factors:

- The increasing production of English manuscripts resulted in an increase in the number of scribes, resulting in more heterogeneous practices.
- The lack of a written standard for the English language until the fifteenth century, which meant that scribes felt no compunction to preserve the spellings of their exemplar in favour of their own spontaneous dialectal forms.
- The development of cursive bookhands that make it difficult to copy the text one letter at a time and encourage the scribe to work “to his own (perhaps silent) dictation”, leading to dialect translation.

Thus, contrary to the general assumptions of textual critical theory,<sup>43</sup> medieval scribes copying vernacular texts did not always (if ever) see it as their duty to produce an exact facsimile-like copy of their source manuscript, as Foulet and Speer (1979: 41) point out. Especially in cases where the exemplar of the scribe was in a different dialect from his own, he could, as McIntosh, Samuels and Benskin (LALME: 1:13) argue, either 1) leave the language more or less unchanged, 2) convert it into his own language variant, or 3) do something in between these extremes, the two latter options appearing to be much more common than the first one.<sup>44</sup> This means that Middle English scribes could be more accurately described as editors of the text they copied, the copy itself being in essence a new edition of the text.

This kind of ‘editorial copying’ was not limited to the linguistic forms of the text. For example Nichols (1986) has argued that the copying of a manuscript text always involves a “mimetic intervention” as the scribe supplants his predecessors, changing words or the narrative order, leaving out or shortening some sections and adding new material to others, bringing about changes that can often be more properly seen as improvements by an intelligent copyist rather than corruptions by a careless one (8). Machan (1994) has observed the same phenomenon, noting that even though scribes were often instructed to copy their exemplars exactly, there is “ample evidence of scribes and correctors rectifying their texts” (Machan 1994: 171). This means, as Caie (2008: 12) has pointed out, that the Middle English textual object was perpetually mutating and evolving as it passed from scribe to scribe, changes being introduced not only by mistakes made by the copyist but also deliberately when the scribe had access to another witness of the work or thought he could improve the text. Especially in the case of popular literature like verse romances—and possibly also utilitarian texts like recipes—which had no significant *authority* behind them, “each act of copying was to a large extent an act of recomposition, not an episode in a process of decomposition from an ideal form”, which was “performed at a level of intellectual and imaginative engagement not inferior to and little different from the putative original act of composition” (Pearsall 1985: 101).<sup>45</sup>

<sup>43</sup> Some textual theorists have, however, recognized these differences and taken a different view of manuscript variation in vernacular texts. For example (McGann 1992) has pointed out that the influence of the variable environment of manuscript reproduction “upon the author’s work in the literary production is by no means always an alien or contaminating influence” (103).

<sup>44</sup> As Benskin and Laing (1981) observe, this ‘dialectal translation’ can operate on different levels, including a) spelling, b) morphology, c) syntax, and d) lexis.

<sup>45</sup> Even in the case of official documents, scribes were not very concerned with following their exemplars exactly, as what mattered both to the compilers of statute books and chroniclers was its gist,

For this reason, as Embree and Urquhart (1987) have argued, the traditional editorial concepts of textual transmission are insufficient and unsuitable for vernacular works whose scribe in essence becomes a co-author by assuming the licence to rewrite the contents of the work “according to his own tastes and biases” (53). This kind of editorial copying, which may be assumed to be more common in texts copied primarily for their information content (Benskin and Laing 1981: 92), plays havoc with the central assumptions of traditional textual criticism (discussed in subsection 3.1.1), leading Greetham (1998) to quite justifiably observe that “the tree organisation of textuality has begun to lose its efficacy as a presiding figure, to be replaced by the rhizome structure (or grass) of Deleuze and Guattari” (299).

This ‘editorial’ approach to textual objects by medieval scribes has been seen to indicate various things in different contexts and by different scholars. While Nichols (1986: 8) sees the scribal reworking of literary manuscript texts as a possible result of changes in fashion or aesthetic taste in the period between the original composition and the copying, and Carlquist (2004) as “a matter of reception, modernisation, or a change in the original text’s use” (112), Grund (2006: 108) sees the substantial variation in the different manuscript versions of the *Mirror of Lights* as a result of knowledgeable scribes or practitioners reworking the text in accordance with their own experience, and thus as an indication of its status as a practical handbook. Similarly, Pearsall (1985), sees the medieval copyist as very much like the modern critical editor, seeing their ‘editorialising activity’ as “evidence of the operation of taste and critical discrimination” (103), identifying areas of the text that were perceived by contemporaries as being unstable or somehow corrupt and thus in need of editorial intervention.

In discussing variation in textual transmission and its implications for editing, it should be kept in mind that the perception of the modern editor is not that of a medieval reader or copyist. This observation places Cerquiglini’s theorization of the medieval text as an “infinitely generating *texte*” (Hanna 1992: 121) on somewhat thin ice, as it presupposes all of the different versions of the text existing concurrently in the social reading community, while in reality they were separated widely in both space and time, individual readers only having access to a limited number of them. Although he admits that medieval works were not necessarily copied from a single exemplar,<sup>46</sup> Hanna (1992) has argued that since no medieval work can be assumed to have been available to everyone everywhere, the access of medieval copyists and readers to the text was always more limited than that of the modern editor, and from the point of view of the medieval audience, “the work is simply whatever is known to the individual patron and producer - which is not the whole tradition as known to the modern scholar” (120).

Finally, if we examine the practice of scribal copying in terms of the layered structure of the textual object described in section 2.1, we can see that textual objects can be copied on different levels: verbatim as a new *document* containing the

---

not its exact words (Clanchy 1993: 265).

<sup>46</sup> Machan (1994: 172) reports one case, found in San Marino, Huntington Library MS HM 114, where a vernacular scribe had collated several copies of the same work in order to improve the quality of his text, although somewhat inconsistently, and Hamel (1998) argues that a medieval text based on a single source is not the norm but in fact “implies a writer in somewhat limited circumstances with perhaps a limited education (since compilation seems to have been a routine technique of clarification and amplification) - or, [...] simply a writer with a limited purpose” (215).



same *text*, with altered orthography as a new *text* representing the same *version*, or with substantial alterations as new *version* of the same *work*. The only level that is not transmissible through copying is the *document*, as the transmission of texts by definition divorces them from documents: texts can only be copied or edited “by being lifted off one document and inscribed upon another” (Gabler 2007: 198). From the point of view of editing—in itself a continuation of the textual transmission of a work—this means that even a documentary approach to editing cannot avoid transposing the text from one document to another. Just as the *text*, also the *visual paratext* can be either transmitted as a facsimile copy or “changed, adapted, or deleted” by the scribe, with potentially significant effects on “the character of the work a reader encountered” (Machan 1994: 165).

This process of textual transmission, consisting of the decoding of information from a source document and re-encoding it into a target one, involving a variable number of steps depending on the level the scribe chooses to perform it on, is represented graphically in *Figure 2.3*. It should be noted the choice of the level on which the transmission occurs is not made anew only at each successive generation of copying, but also *at every successive textual segment*: the scribe might well decide to copy some parts of his source document verbatim on the level of *text*—producing an exact copy subject only to errors caused by his misreading or miscopying something—and some parts on the level of the work, providing in essence a paraphrase of the general idea presented by the source text but in a significantly condensed or elaborated form. Furthermore, as was pointed out above, the scribe was not limited to a single source document. Not only could he copy different parts of the text from different source documents or even different parts of the work from different versions—producing essentially the equivalent of a critical edition—but also include material from entirely different works or even original material of his own, serving as a *compiler* or a *co-originator*.<sup>47</sup> This kind of ‘editorial’ copying—represented by the dashed lines in *Figure 2.3*—would always produce a new *version*, regardless of the level of copying employed.

From the point of view of historical linguistics, the most significant implication of the multilayered genesis and transmission of surviving medieval documents lies in its interplay with the concept of *context* discussed in section 2.2 above. The fact that each copying of a manuscript text amounts to a new process of textual creation means that each copying also instantiates a new context which inevitably influences the new *version* and *text* that is produced and encoded in the *document* created by the act of copying. Thus each surviving document carries within it the remnants of not just one but a number of production contexts, resulting in a linguistically complex and layered textual object.<sup>49</sup>

<sup>47</sup> Like Carroll (2003: 156), I do not use the term *compiler* in its technical sense of *compilatio* or a systematic anthology of established *auctores* as described by Parkes and Doyle (1978: 190) and Parkes (1991a), but rather in its broader sense of an assemblage of different kinds of material into a composite text (as used by Parkes 1991b: 292).

<sup>48</sup> The dashed lines represent cases where new material is added to the textual object otherwise copied on the level of the *text* or the *version*, producing a new *version* of the work.

<sup>49</sup> Benskin and Laing (1981) have examined some of the linguistic consequences of this process and the resultant hybridity from the point of view of dialectology, but unfortunately they are not always taken into account in historical linguistic studies.

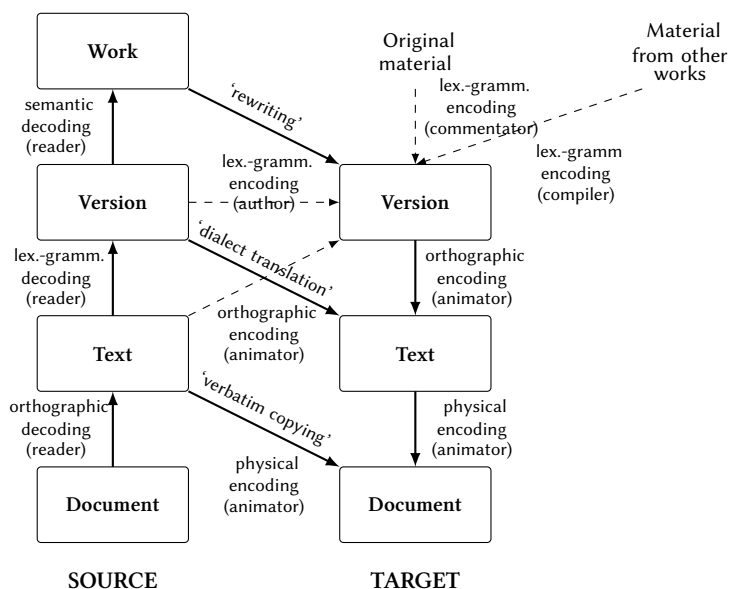


Figure 2.3: The different levels of manuscript copying, including the processes and roles involved in them.<sup>48</sup>

## 2.4 Transcription of text

But it seems to me, sometimes, that readers and editors may be seen as well, *even as they are readers and editors*, as authors and writers. And it also seems to me that authors and writers may be seen as well, *even as they are authors and writers*, as readers and editors. I am not ‘free’ with respect to this text I am writing. Even as I write it I am reading it as if I were in another time and place—as if I were here and now, in fact—and my text, my ‘textualité’, is constrained and determined by a future which at all points impinges upon my present text. This is to be in the textual condition. (McGann 1991: 95)

McGann’s description of what he calls the ‘textual condition’ highlights the fact that in transcribing the text of a medieval document, the modern editor is in exactly the same position as the medieval scribes copying the manuscripts. As Pidd and Stubbs (1997) have observed, this means that any modern transcription—and the edition based on it—is simultaneously “an undertaking to try and comprehend the [scribal] tradition” (58-9) of a manuscript text and a continuation of that scribal tradition, adding yet another layer of scribal transmission to the textual object. We have already established that the physical *document* is our only source of evidence of the text it encodes and should therefore be the main focus of our attention. Considering that the document is also the only aspect of the textual object

that cannot be transmitted—either by the medieval scribe or the modern editor—it becomes obvious that transcription—like scribal copying—is far from being the simple act of reproduction it is sometimes made out to be in discussions of editorial theory.<sup>50</sup>

Traditionally, the aim of ‘literal transcription’ has been taken to be the reproduction of ‘the text’ of the original document, or as Vander Meulen and Tanselle (1999) formulated it, “to report—insofar as typography allows—precisely what the textual inscription of a manuscript consists of” (201). The problem with this, of course, is the fact that different editors have very different and often somewhat hazy views on what this “textual inscription” actually consists. As Renear (2001) has somewhat whimsically put it, the extraction of ‘the text’ from a document “is a simple matter without theoretical significance, only for those who have never tried it” (33). Thus, as Dahlström (2009) has rather forcibly argued, an editor—like a medieval copyist—can never simply *reproduce* the original document but rather “engages in creating a new, target document, ‘similar’ but all the same derivative from the departure material” (34). Thus, even the most “diplomatic transcription is already a distinct abstraction from the document” (Gabler 2007: 204).

In the case of modern editing, where the aim is not only to produce a copy, but to produce a copy in a *different medium*—either a printed book or a digital document—there are further complications. The editing of manuscript texts “always has been and always will be a transformation from something into something different” (Rehbein 2008: 1). Thus, as Robinson and Solopova (2006) have observed, the digital transcription of a manuscript text “cannot be regarded as an act of substitution, but as a series of acts of translation from one semiotic system (that of the primary source) to another semiotic system (that of the computer)” (2).<sup>51</sup> Differences in the expressive capabilities of the source and target media mean that transcription always involves a process of *selection* Sperberg-McQueen (2009: 31).<sup>52</sup>

In the [media] translation process, certain features of the work are preserved that can be carved into the flesh of the new medium and can be expressed by its architecture and the language of its web of signs, while others are treated as noise, obscuring the substantive signals.  
(Dahlström 2009: 33)

Pierazzo (2011: 464) sees this selective nature of transcription to have two

<sup>50</sup> For example Renear (2001) and Haugen (2004) refer in critical tones to the consensus within editorial theory and methodology that “literal transcription is a relatively simple matter” (Renear 2001: 32) and “the transcription of a primary source is often seen as a rather simple and unambitious activity” (Haugen 2004: 73).

<sup>51</sup> In editing modern printed texts, however, the similarity of the two semiotic systems means that it may be possible to produce a transcription “that a scholar for most purposes can use [...] as a substitute for the source” (as Bøe, Jørgensen and Taugbøl (2004: 58) have argued for the writings of Henrik Ibsen).

<sup>52</sup> As was already pointed out, editors have quite varying views on what are the features of the textual object that should be preserved: “some respect typographical or scribal errors, some correct spelling, some render punctuation conventionally, or at least consistently, some alter paragraphing” (Shillingsburg 1986: 26), and some—like the present author—propose to include “not only the linguistic text expressed in linear sequences of alphanumeric characters along with punctuation, but also the accidental textual particulars (expressed by typography and other visual markers) that McGann labels bibliographic codes” (Dahlström 2009: 33).

important consequences:

- 1) “no transcription, however accurate, will ever be able to represent entirely (i.e. faithfully) the source document” (Pierazzo 2011: 464), as some of the infinite number of characteristics of the manuscript are always lost, (as was recognized even by Vander Meulen and Tanselle (1999: 201)); and
- 2) “any transcription represents an interpretation and not a mechanically complete record of what is on the page” (Pierazzo 2011: 465), the process of transcription being “a systematic program of selective alteration, coupled with selective preservation of information” (Huitfeldt and Sperberg-McQueen 2004: 302).

Contrary to the traditional view, transcription cannot thus aim at ‘true representation’ of the original document, but rather at a representation that is appropriate for some purpose and audience (Pichler 1995b: 690). The purpose and audience of a transcription thus define, to a large extent, both *what* is represented and *how* it is represented. The selective nature of transcription has led some scholars like Robins (2004) to argue that the term *representation* is in fact not an accurate description of the relationship between the textual object and its transcription:

Any claim that an edition *represents* a text is grievously overstated, for editions are very selective about what features of texts they attend to, and they do not in fact ‘re-present’ any textual object in its fullness. Rather, an edition *models* some of the determinations of a text. Like any model, it raises one set of features to visibility by excluding others in order to serve a heuristic purpose. (Robins 2004: 146)

In their formal conceptualization of the process of transcription, Huitfeldt and Sperberg-McQueen (2004: 296-9) have characterized this process of *modelling* as being constituted by 1) the identification of a sequence of physical *marks* in the source document that is considered significant, 2) the identification of these marks as *tokens* of abstract *types* from a pre-determined typology, and 3) the production of a target document that contains a new sequence of tokens that instantiates the same sequence of types as the source document.<sup>53</sup> As Huitfeldt and Sperberg-McQueen (2004) point out, this has certain consequences for the nature of transcription. First of all, in addition to information left out as ‘irrelevant’ by the transcriber, the transcription process also loses some information by virtue of being a process of *abstraction*: the mapping of tokens to types and back to tokens means that for example any graphetic detail about the original token is lost, and “the only salient information about the token in the exemplar is which type it maps to” (301). For this reason, the repertoire of types to which tokens are mapped is of prime importance, as its size—i.e. the number of distinctions it makes—determines the amount of information preserved by the transcription:<sup>54</sup>

Sign systems have a discrete character. To set up a sign in place of a thing therefore means, at the same time, to standardize it, to choose

<sup>53</sup> In terms of the model, it is irrelevant whether these tokens and types are characters, words, sentences or texts, but in the present discussion, they are assumed to be individual characters (Huitfeldt and Sperberg-McQueen 2004: 298).

<sup>54</sup> The distinctions made in transcribing letterforms in the present edition are discussed in subsection 10.2.2.

an individual member suitable to represent it from a finite number of members of a paradigm. If a series of signs obtained through such a selection is to be the object of scholarly research, then the suitability must be guaranteed by establishing that the entire operation proceeds according to previously given, logically consistent, and generally recognized rules. (Červenka 1995: 59)

The fact that transcription is based on “the recognition of unitary tokens and their mapping into equivalence classes (types)” (Huitfeldt and Sperberg-McQueen 2004: 302) means that it is always a process of *digitization*, as the continuous (or *analogue*) phenomena of the original document are converted into discrete (or *digital*) data through a process of categorization.

As was already mentioned above, this process of categorization is an analytical activity that always “involves the application of competent intelligence” (301). This means that the complex operation of selectively decoding data considered relevant from the original document and re-encoding it into a new document in some other medium (printed or electronic document)—essentially a process of what is known in information technology as ‘lossy compression’—is always “fundamentally incomplete and fundamentally interpretative” (Robinson and Solopova 2006: 2) in nature and always involves alterations which are interpretative and irreversible: no two scholars, even when given the same transcriptional criteria, are likely to produce identical transcriptions of the original, and it is impossible to reconstruct the original document on the basis of the transcription (Pierazzo 2011: 465).<sup>55</sup> Huitfeldt and Sperberg-McQueen (2004) have divided the sources of disagreement in the interpretive process into two basic categories. First of all, transcribers may disagree whether a particular physical phenomenon (i.e. a pen stroke) in the document should be considered a significant token (e.g. a letter, as opposed to an accidental mark of the pen) (306), and second, they may, while agreeing that a particular mark is a token, “disagree about how to read a particular token in the exemplar” (303), i.e. to what type it should be assigned to.

This process of interpretation and categorization is significantly complicated by the fact that it is not based exclusively (or even primarily) on the objective graphic qualities of the mark on the page, but rather on the subjective interpretation of the textual context. For example, while the sequences of strokes constituting the word *in* in most medieval hands are graphically identical to those constituting the letter ⟨m⟩, they are interpreted and encoded differently by the transcriber based on the context: “An *i* is an *i* not because it is a stroke with a dot over it. An *i* is an *i* because we all agree that it is an *i*” (Robinson 2009: 43). Thus, the remark of Sir Walter Greg that bibliographers are not interested in the meaning of the letters on the page but treat them simply as arbitrary signs is simply fallacious, as Eggert (1994) has pointed out: in order to perform any bibliographical operations on “marks on the page” they need to be “raised onto a textual plane, be given a meaning, if only a provisional one, before they can be interpreted” (1–2). The transcriber’s job is thus precisely that of interpreting the “marks on the page” and abstracting them into textual content.

<sup>55</sup> In information technology, this could be seen analogous to two different implementations (‘transcriber’) of a technical specification (‘transcriptional criteria’) producing two results that, while fulfilling the general specification, differ in numerous details not governed by the specification.

Thus it is painfully obvious that transcription “is a much more complex process than ‘to represent the original manuscripts as correctly as possible’”, and involves selection and “a range of different interpretational activities” (Pichler 1995b: 690). The principal reason for this is that the *text* to be transcribed is not a ‘thing’, but a *process*—the construction of *meaning* based on a linguistic interpretation of physical marks in a document and various kinds of cultural and situational knowledge about its functions and significances. The medieval manuscripts are not and never were “merely transmitters of abstracted verbal text”, but “sites of textual transaction and of social performance” (Eggert 2009: 72):

It is, ultimately, a metaphysical transformation that will be required to domesticate the difficult idea that a representation is not a representation of pre-existing entity, or that the integrity of a text - these words in this order - matters precisely because it is the condition for *doing* the work of textuality, a work of humane reasoning about relations that can be captured in words. (Searle 2004: 19)

What transcription then involves, is the conferral of textual and paratextual meaning to certain aspects of the physical *document*, and their classification and interpretation as abstract meaning-bearing units which can then be encoded according to certain conventions into a new document. Transcription is thus not an objective description of pre-existing reality, but rather “an argumentative statement on the constitutive components of the departure document” (Dahlström 2009: 41), a material encoding of a textual—or ideational—interpretation of the departure document, necessarily limited both by the values and textual views of the transcriber and by the fact “that neither printed nor electronic formats can begin to represent the wealth of data provided on a manuscript page” (Parker 2006: 207).

## 2.5 Conclusion

The study of genre, the author’s other works, biography, cultural history, the history of ideas, all these are understood to extend our awareness of the contexts within which texts create and convey meaning. Even the physical embodiments of texts, the books themselves as paper, ink, and bindings, influence interpretation.

(Shillingsburg 1991: 28)

What has become clear in the preceding discussion is that ‘texts’ are in fact multilayered, multiply contextualized and constantly changing processes of meaning-creation, and any editorial activity must thus by necessity be thoroughly subjective, interpretive and selective. However, the fact that an edition can never capture the totality of the original document should not be seen as much as a methodological shortcoming but rather as a historical inevitability. As Machan (1994: 78) points out, we can have no access to the text as it existed in its own time. In reading a medieval document, our experience can never be commensurate to that of a

medieval reader, who was not reading a ‘foreign’ *medieval* document but rather a thoroughly familiar contemporary one, produced and used in a cultural context with which he was intimately familiar, but of which we know practically nothing beyond that which has happened to survive through those very documents we are studying. Being constructed in a modern context, the texts we—as editors—decode from medieval documents are by necessity different from the medieval text, and our conception of the work embodied by these texts may be completely different from the medieval work.

The view of the manuscript as a textual object and the processes of its transmission described above do not bode well for the aim of traditional textual criticism, defined by Maas (1958: 1) as “to produce a text as close as possible to the original”, as we will see in section 3.1. Taken together, they lead to the observation that in any edition that we prepare, not only do “the words on the modern printed page bear an indeterminate relation to what the author originally wrote”, but “it may be impossible ever to discover what the author originally wrote - and worst of all, it may be that conventions of text production in the medieval period render the concept of an original, authorial text simply irrelevant” (Minnis and Brewer 1992: x). Fortunately, this does not mean that we could or should not produce editions of medieval textual objects. It merely means that we will need to consider carefully the purposes for which we intend to use our editions for and the way in which we go about producing them.

This thesis and the edition it contains—and especially this chapter of it—is an attempt to heed the warning of Blake (1992a: 38) that until we, as editors, start to pay more attention to the implications of medieval text production and transmission for the editorial process and our methods, we will continue to “pay lip service to modern textual studies while simply reproducing what has become standardized in editions” (38). More specifically, it tries to acknowledge the multifarious nature of medieval textual objects and build its theoretical and practical methodology around Blake’s observation that we may have to accept different versions of texts with not only textual but also structural variation if we wish to offer a useful viewpoint on medieval texts.





## Chapter 3

# Editing historical manuscripts

As was already out in the preceding chapter, the preparation of a scholarly edition of a manuscript text is not fundamentally different from any other step in its textual transmission in that it involves the encoding of a *work*, a *version* or a *text* decoded from one or more earlier *documents* into a new document, traditionally a printed book. As such it is always “a positive construction in its own right”, and the fact that its intended function and target audience are by definition quite different from those of the original document, the process of editing always involves “an act of reimagining and redefining a text’s audience(s) and its ways of interacting with those audience(s)” (McGann 1991: 66), or an act of ‘functional translation’. This process of translation can be seen to proceed from the *departure document* to the *target document* in several phases, outlined by Dahlström (2009: 32) as: 1) scrutinizing a document; 2) establishing what particulars in the document are to be considered substantive elements of the work we suppose the document to contain; 3) representing these substantive elements in some form, using a medium either similar to or different from that of the original; and 4) creating a new *derivative* document that “purports to be a remake of the departure document, and to some extent, of the work which the latter contained” (32). Because this translation process—just like any other process of manuscript transmission—takes place in a specific context by a specific individual, it always includes an element of interpretation and can never be purely objective.<sup>1</sup>

Although the principal aim of scholarly editing is sometimes considered to be making “available for scholarly use works not ordinarily available or available only in corrupt or inadequate forms” (Shillingsburg 1986: 2), scholarly editions should not be seen “as neutral prolongers of the life of the works and documents” (Dahlström 2009: 44), but rather as analytical tools for their study. This re-pur-

---

<sup>1</sup> Dahlström (2009: 32) has listed some specific contextual parameters that he sees as influencing this process of *media translation* and consequently the shape taken by the scholarly edition: 1) the socio-cognitive, psychological, linguistic particulars of the *individual(s)* responsible for carrying out the translation; 2) socio-cultural and socio-technical particulars of the *situation* in which the translation takes place (e.g. culture and tradition, purpose, specific audience, media environment); 3) the material and technical particulars of the departure and target *media* (such as supporting matter, longevity, compatibility, document architecture); 4) physical or symbolic *tools* at use in the process (such as practices and techniques, software, platforms, requirements, regulations and rules), and so on.

posing of the edited works and documents, as well as the radical break between the original and the editorial contexts (on all three levels outlined in section 2.2), is what differentiates the scholarly editorial process from a scribal one.

The intermediate position of the editor between the document and its production context on the one hand and the edition and its reception context on the other has led Gumbrecht (1998: 243) to see the editorial role as comprising of a hypothetical author-role and at least one—possibly several—reader-roles: the editor must simultaneously situate herself “in a historical relation to the work’s transmissions” and “in an immediate relation to contemporary cultural and conceptual goals” (McGann 1991: 47). This dual task means that in selecting an editorial methodology, the editor has to be aware of both the object (i.e. the properties of the source document) and goal of the edition, or as Zeller (1995a) has put it, “editorial measures should be proper to their object and suitable to their social function” (19). In practice this means that in formulating her editorial principles and practices, the editor must take into account—and explicitly acknowledge—not only the historical production circumstances of the original document(s) and the model of textuality implied by them (see e.g. Gaskell 1978, Gabler, Bornstein and Pierce 1995 and Kline and Holbrook Perdue 2008), but also the purposes for which the edition is intended to be used (see e.g. Gaskell 1978 and Shillingsburg 1986).

Accordingly, the concept of *scholarly edition* has come to cover “a spectrum of variant types ranging from facsimile, diplomatic, synoptic, genetic, critical, variorum editions to large-scale digital archives on compact discs or mounted on the web” (Dahlström 2009: 28). Shillingsburg (1986) has distinguished between what he calls four major “formal orientations” represented by editors: *historical*, *aesthetic*, *authorial* and *sociological*. These orientations do not represent—or even dictate—a specific editorial approach, but rather “a perspective on forms which leads to the selection of one set of formal requirements over another” (19). The *historical* formal orientation,<sup>2</sup> which “places a high value on the chronology of forms” and avoids the mixing of historically discrete documents, can be seen as the ‘foundational’ orientation, the others representing essentially different grounds on which “deliberate violations of documentary historical forms” can be justified (20).

It should, however, be noted that the orientations defined by Shillingsburg do not represent—or even dictate—particular editorial strategies, but are rather value systems that can be used to support specific editorial approaches. For example the historical orientation can equally well be used to support a strictly diplomatic editions as an edition aiming to purge “nonauthoritative” changes from a text, or an edition representing the author’s “final intentions” (Shillingsburg 1986: 20). However, since the other orientations are fundamentally means “to ‘correct’ historical forms” (21), they always result in “an eclectic text which is as a construct non-historical” (Vanhoutte 2009: 105). These orientations also illustrate the fact that while different types of editions are often characterized by varying degrees of ‘faithfulness’, the concept of fidelity is not an unambiguous one and can be evoked in relation to any of the levels of the textual object outlined in chapter 2 to produce quite different editorial approaches (Edwards 2000: 66).

<sup>2</sup> As Vanhoutte (2009: 105) points out, Shillingsburg renamed this orientation as the “documentary orientation” in the third edition of his text.

Despite the problematic nature of ‘fidelity’, it has often been used as the organizing principle in categorising different editorial approaches. For example Greetham (1987), in the context of editing the *Regement of Princes*, has listed and described seven alternative editorial approaches, listing them in “declining degrees of ‘fidelity’”, starting with the *photographic facsimile* and progressing through the *diplomatic transcript*, the *best-text edition* and the *copy-text edition* to the *genetic edition*, the *Slavic textological edition* and the *social textual edition* (62-9). While the first four types can be placed on this cline reasonably comfortably, provided that we define fidelity in relation to the original document, the last three types of edition cannot really be claimed to be in any sense ‘less faithful’ than the first four, being instead based on entirely different views of the text.

The German editorial tradition has also defined similar classifications, although on slightly different grounds. For example Klaus Kanzog in his *Prolegomena zu einer historisch-kritischen Ausgabe der Werke Heinrich von Kleists* (1970) distinguished between four historical types, summarized by Vanhoutte (2009: 107) from top to bottom as:

- 1) the *archive edition* (‘Archiv-Ausgabe’), which “provides the exact documentation in a useful system of all witnesses and textual phenomena including all genetic and transmissional variants”;
- 2) the *historical-critical edition* (‘Historisch-Kritische Ausgabe’), which “assesses this material, provides an established text and orders the variants from the perspective of this text”;
- 3) the *study edition* (‘Studienausgabe’), which “provides a scholarly established text, ideally derived from the historical critical edition, together with a commentary section that does not only treat the critical and historical status of the text but also its interpretation”; and
- 4) the *reading edition* (‘Leseausgabe’), which “is intended for the general reading public and can, depending on the choice of the reading text, the editorial principles and the facultative commentary, be a critical or a non-critical edition”.

Although these edition types can also be seen in terms of ‘fidelity’, it is important to note that this classification is primarily based on the *purpose* of the edition, differences in the degree of fidelity being instead a consequence of the editorial methods adopted to answer the needs of that purpose. In this sense Kanzog’s classification is similar to Shillingsburg’s, bearing only an indirect relationship to the actual editorial methods employed by the editor. Both of these classifications, as well as the failure of Greetham’s classification to locate the different edition types onto a single cline, demonstrate that different editorial approaches should not be seen as competing solutions to a single problem, but rather—as Gumbrecht (1998: 248) has pointed out—more like craft guilds that have their own traditions and practices, aimed at different ends. Thus the choice between competing editorial approaches becomes not a question of choosing the best or most effective one, but rather choosing the best or most effective one *for a particular purpose*.

In addition to these kinds of hierarchical typologies, several scholars have also noted a simpler, more fundamental division of editions into two types based on two competing viewpoints to the *textual object*. The first of these can be characterized as a striving for “simple, accurate representation” (Modiano, Searle and

Shillingsburg 2004: xiv) or the attempt of archivist editors to preserve cultural objects (Shillingsburg 1999: 60), and the second as the creation of “critical, interpretive added value” (Modiano, Searle and Shillingsburg 2004: xiv) or the attempt of literary editors to rectify “the vicissitudes of history in a fallen world” (Shillingsburg 1999: 60). Gleßgen and Lebensanft (1997: 10) have defined these two competing principles as fidelity to “der (*konkreten*) Faktizität des überlieferten Dokuments” on the one hand and fidelity to “der (*abstrakten*) Idealität eines stets unerreichbaren ‘Originals’” on the other (Gleßgen and Lebensanft 1997: 10). Bodard and Garcés (2009: 92) have translated these principles as “the concrete and factual materiality of the extant documents” on the one hand and “the principle of the ideal and abstract notion of a reconstructed archetype” and outline a continuum of scholarly edition types between these extremes, starting from the artefact itself and moving on to “a variety of surrogates (drawn, photographed, scanned)”, then on to “‘diplomatic’, normalized, synoptic and ‘best-text’ editions” and finally to “the full-on historical critical edition, complete with apparatus variorum”. While the resulting typology is quite similar to that of Greetham mentioned above, the crucial difference lies in the bidirectionality of the cline; instead of being defined as a single-ended cline of fidelity, it is here defined as a dual-ended one, at the one end of which lies maximum fidelity to a physical *document* and at the other maximum fidelity to an abstract *work*.

The division of this cline of edition types into two groups somewhere down the middle corresponds to the fundamental division, suggested by several scholars (see e.g. Foulet and Speer 1979: 42, Shillingsburg 1986: 52, and Williams and Abbott 1999: 71) into *documentary* (or *diplomatic*) and *critical* editions, the former presenting “a text identical to that in a historical document” and the latter “a text that mixes material from two or more versions according to some critical dictum” (Shillingsburg 1986: 52).<sup>3</sup> Over the last two decades, the field of scholarly editing in general has seen a movement away from critical editions and towards ones with “evidentiary value”, characterised by David Greetham as ‘documentalism’ (Kline and Holbrook Perdue 2008: 24). This same phenomenon, which has—whether coincidentally or not—proceeded hand in hand with the increasing digitalization of scholarly editing (see chapter 4) has also been observed by Shillingsburg (2004):

In recent years it has become far more common for editors to provide access to historical texts than for editors to provide established texts that render textual problems and complexities transparent. It is more likely that an editor will speak of placing the text in its context than that the editor will claim to have established the definitive edition for a literary work. It is common now for editors to refer to their work as acts of criticism or as choices among several viable editorial acts.  
(Shillingsburg 2004: 414)

The two main sections of this chapter will introduce and evaluate the strengths and weaknesses of these two basic approaches in terms of the aim of this thesis,

<sup>3</sup> Shillingsburg (1986: 52) also includes in his division an intermediate type, an edition whose text represents “a historically identifiable ‘version’”, which is based on all the documents representing the same *version* of the work, thus implying more normalization than one based on a single *document* as well as the omission of all document-specific features.

namely the production of *digital editions of medieval vernacular utilitarian manuscripts* for the purposes of *corpus linguistics*.

### 3.1 Critical editing and textual criticism

A critical edition is a kind of text which does not seek to reproduce a particular past text, but rather to reconstitute for the reader, in a single text, the entire history of the work as it has emerged into the present. (McGann 1992: 93)

The discipline of critical editing was born in the context of classical and biblical studies to answer the practical need “to recover, or approximate by historical reconstruction, the lost original works of ancient authors” (McGann 1992: 23), inspired by the contemporary comparative linguists’ attempts at reconstructing the Indo-European *Ursprache* by comparing surviving language-forms, and based on the “systematic collation of all the relevant texts of the work in question” (23). In the 19<sup>th</sup> century it was introduced to the study of vernacular literatures, forming the basis of what we now consider modern philology—“the professional study of national scriptures” (McGann 1985: 184-5). Despite its rather specific original context, the critical editing has achieved a surprisingly hegemonic status in the editing of various types of texts over the 20<sup>th</sup> century. Its principal objective—“to produce a text as close as possible to the original” (Maas 1958: 1)—has become virtually a given, discussions of editorial theory mostly focusing on *how* one should go about achieving this objective. This has meant that it has frequently been presented as a universally appropriate method, suitable even for Middle English texts—at least those that have been considered worth editing in the first place (Machan 1994: 60).

However, it is important to remember that its humanist focus on the original *authorial* text, based on the Renaissance valorization of the individual self (Machan 1994: 14, 179), and its aim of producing “a text as close as possible to the original” (Maas 1958: 1) are intimately tied with their original context of biblical and classical scholarship (McGann 1985: 186), and thus based on some very specific presuppositions both about the nature of textuality and the purposes of edited texts. In order to provide a context for the discussion of the theoretical and practical orientation of the present edition and to justify its radical departure from the established methodology of critical editing, this section will first provide a very brief overview of the theory and practice of critical editing and then proceed to problematize them in the light of medieval textuality, the demands of linguistic research, and the possibilities offered by digital editing.

### 3.1.1 Theory and methodology of critical editing

The theoretical basis of the methodology of critical editing lies in *textual criticism*, a subfield of *literary criticism* concerned with the identification and removal of transmission errors in the surviving documentary forms of literary texts—a historical solution for dealing with “transmissional noise” (Dahlström 2009: 32). According to the theory of textual criticism, the elimination of “textual contaminants” and the “interfering scribal and typographical presence” (40) will result in the restoration of the authorial text, much like an old painting is restored by stripping away its “overlayers of varnish and smoke discolouration [...] to reveal the true object” (Eggert 1991: 61). This idea of ‘reconstruction’ or ‘recovery’ (as well as the vocabulary commonly used to describe the relationships between variant versions of a work) is based on the concept of *degeneration*, positing a single act of composition by an original author—embodying *authorial intention*—which is followed by a series of scribal copies, each adding errors that lead to its corruption. As several scholars (see e.g. Embree and Urquhart 1987: 52 and Allen 1987: 21) have pointed out, this view places high demands on the original author and his act of composition:

La thèse de la copie comme dégénérescence présuppose un original sans faute : l’auteur n’a pas droit au lapsus. De même, l’idée de la dégradation langagière implique une origine impeccable : l’auteur n’a pas droit, non plus, à l’incorrection, à l’à-peu-près, voire à la diversité de sa parlure. La philologie, ce faisant, s’adjoint de façon subreptice une théorie littéraire qui est celle du génie.<sup>4</sup> (Cerquiglini 1989: 90)

These concepts of authorial intention and its recovery through the elimination of scribal error, originally formulated in the context of classical editing, were gradually accepted very much as a given, and dominated Anglo-American editorial theory in various guises well into the 1990s, when their hegemony finally started to break down as a result of challenges by the *social textual criticism* of Jerome J. McGann and D. F. McKenzie on the one hand,<sup>5</sup> and more document-oriented and linguistically aware approaches on the other (see sections 3.2 and 4.3). While the reconstruction of an archetype resembling the authorial intention as closely as possible has been widely accepted as the core task of critical editing, there is less agreement on whether this reconstruction is possible and how it should be achieved. The first question divides critical editorial approaches in two: those which believe in the possibility of reconstruction and those which do not. The approaches answering this in the affirmative—*eclectic editing* and *Lachmanian recension*—can be considered ‘reconstructive’ approaches, because they take

<sup>4</sup> “The theory that the copy represents degeneration presupposes a flawless original; the author has no right to any lapsus. Similarly, the idea that language becomes degraded implies an impeccable origin: the author has no right to bad language either, or to dreadful puns, or, indeed, to the diversity of his way of speaking. By subscribing to this, philology surreptitiously annexed a literary theory—the theory of the genius.” (Translated in Cerquiglini 1999: 61.)

<sup>5</sup> The effect of McGann’s ‘social textual criticism’ has been to direct attention away from the author and the concept of ‘authenticity’, “focusing it instead on forms and structures deriving from the culture as a whole and from language itself” (Shillingsburg 1986: 11). This approach has—not unpredictably—taken root especially in the editing of texts that have not been endowed with a strong *author function* such as medieval romances, and of course non-literary, especially practical writing.

as their task the synthesis of the surviving manuscript tradition into an ‘eclectic’ text that does not correspond to any existing document, but is an aggregate of those readings that the editor judges most likely to be authorial. The differences between Lachmannian recension and eclecticism are mainly methodological: whereas eclectic editors rely solely on the informed judgement of the editor in selecting authorial readings, frequently leading to “the establishment of a number of readings unsupported by any authority” (Edwards 1987: 47), the Lachmannian ‘scientific’ method bases its choices on a ‘family tree’ or *stemma codicum* of the surviving manuscripts, established on the basis of shared errors, restricting itself to readings found in one or more surviving manuscripts.

In contrast, the more conservative approach of *best-text editing* can be termed ‘preservative’ in nature, as it rejects the eclectic combination of manuscript readings and instead bases the edition on a single manuscript deemed by the editor to be as close as possible to the authorial original (Bodard and Garcés 2009: 95), resulting in “a text with some historical basis, not one manufactured eclectically by a modern scholar” (Fahy 2004: 403), with emendation restricted “to the correction of its obvious errors” (Williams and Abbott 1999: 75). In 1950–1951, in his seminal article *The Rationale of the Copy-Text*, Sir Walter W. Greg (1875–1959)—originally a follower of Bédier—took the best-text approach, which he saw as too conservative and limiting for editing Shakespeare and other Renaissance drama, and developed into what became known as the *copy-text* approach by dividing the text into what he called *substantives* and *accidentals*.<sup>6</sup> Based on his observation that compositors at early modern printing shops tended to aim at a faithful reproduction of the *substantive* readings of their copy while following “their own habits or inclination” (Greg 1950–1951: 22) with regard to accidentals, he proposed to treat these two aspects of the text differently in the edition. His solution was to follow the *copy-text* for the accidentals, but to apply to the substantives “a technique of controlled eclecticism whereby the editor, in the light of all the evidence, emends the copy-text by substituting readings from another text or by supplying new ones by himself” in those places where he believes the copy-text to differ from the author’s intention (Gaskell 1978: 4–5), essentially creating a hybrid of best-text and eclectic editing. Although the method was originally developed strictly in the context of printed English Renaissance drama—its rationale being based on the specific practices of early printing industry—the theory of the copy-text was subsequently presented as a general editorial strategy—or rather the “supreme” editorial strategy—by Fredson Bowers (1905–1991) and Thomas Tanselle (1934–),<sup>7</sup> and ended up dominating discussions of editorial theory and the editing of modern literature for much of the late 20<sup>th</sup> century.<sup>8</sup>

<sup>6</sup> “[A] distinction between the significant, or as I shall call them ‘substantive’, readings of the text, those namely that affect the author’s meaning or the essence of his expression, and others, such in general as spelling, punctuation, word-division, and the like, affecting mainly its formal presentation, which may be regarded as the accidents, or as I shall call them ‘accidentals’, of the text.” (Greg 1950–1951: 21)

<sup>7</sup> They saw it as “the most workable editorial principle yet contrived to produce a critical text that is authoritative in the maximum of its details” regardless of the literary genre or period (Bowers 1972: 86).

<sup>8</sup> For example practically all of the scholarly editions approved by the Modern Language Association’s Committee for Scholarly Editions have been prepared according to its criteria and assumptions (Cohen and Jackson 1991: 106).

While all of the four aforementioned approaches have had their advocates in the Anglo-American editorial tradition of the 20<sup>th</sup> century,<sup>9</sup> there has been remarkably little theoretical reflection on their application to editing Middle English texts. Instead, Middle English editors have in general “paid rather scant attention to the theoretical advances in editorial procedures” (Blake 1998: 76), and in the main settled for what appears to be a quite widely varying and for the most part scantily documented set of editorial practices influenced—to varying degrees—by the different theories of critical editing mentioned above, depending on their specific aims and textual situation.

In terms of the textual model introduced in chapter 2, the aim of all these varieties of critical editing can be defined as trying to reconstruct the underlying *work* on the basis of all or some of the surviving documents and the texts they contain, and then to reconstruct for it a *textual realization* that is as close as possible to the one set down by the author. In terms of textual transmission, as represented in Figure 2.3, it thus takes the maximally long route from the surviving document(s), through the texts contained in them and the versions represented by these, to discovering the abstract work intended by the author, and then tries to reconstruct an authorial version of it, clothing it in linguistic forms that it sees most likely to represent those used by the author, and finally formatting this into a document following the conventions of the modern printed scholarly edition. This extended route of course means that not only the visual paratexts of the original documents (or their ‘bibliographic codes’) get omitted and replaced by those of modern printed texts during the process of abstraction and reconstitution (Eggert 1994: 19), but also the specific textual features of the original manuscript texts often get replaced by new ones reconstructed by the editor.

### Criticism of critical editing

Although still in many ways the ‘default’ mode of editing English literary and even historical texts,<sup>10</sup> the last two or three decades have seen an increasing amount of criticism levelled against the different methods of critical editing. Some of the criticism has focused on the practical or theoretical shortcomings of a specific critical approach without challenging the basic premises of critical editing outlined above, but this chapter will concern itself primarily with problems stemming from those

<sup>9</sup> Bediér’s best-text editing can be seen to have influenced the rather conservative and philologically oriented early editions of the EETS, while the Lachmannian approach was advocated in England mainly by A. E. Housman (1859–1936), a classical editor who savagely attacked best-text editors for shirking from the task of critical judgement (Walsh 2010b: 160), and its most notable product is the monumental work of Manly and Rickert (1940) on *Canterbury Tales*. Eclectic editing, on the other hand saw a major revival in the editions of the *Piers Plowman* A and B versions of the *Piers Plowman* (Kane 1960; Kane and Donaldson 1975) by George Kane (1916–2008) and E. Talbot Donaldson (1910–1987), which are best characterized as arrogantly eclectic reconstructions of an authorial text by the editors and have been both hailed as works of unparalleled editorial genius and reviled as cautionary examples of editorial hubris. Despite its theoretical significance, Greg’s copy-text approach has “not evoked much interest from editors of Middle English works” (Edwards 1987: 46), but has made the distinction between *substantives* and *accidentals* a commonplace even in discussions of Middle English editing.

<sup>10</sup> Considering its shortcomings, it seems rather incredible that as late as 1994, Machan could observe that the tradition of classical textual criticism in the vein of Greg, Bowers and Tanselle “still represents the textual-critical status quo to which Middle English [...] has been accommodated and against which genuine theoretical challenges [...] have recently been articulated” (36).



basic premises themselves. As was already mentioned above, textual criticism has always considered “its objectives, its principles, and, above all, itself as givens” (Machan 1994: 33) and seen its methods as universally applicable across both historical periods and types of literature. This idea of universal applicability is based on a conception of editing as an ‘objective’ endeavour, the ‘establishment’ of texts that other scholars then go on to ‘interpret’, which is of course patently false:

[C]hoices among authoritative forms of the text are in no sense scientific or objective. This is a fact to be acknowledged, not a weakness or flaw in any editorial theory. But nearly all scholarly editors abhor this fact and [...] are always careful to justify as principled and objective the editorial choices they must make. (Shillingsburg 1986: 86)

Instead of taking any editorial method as objective and universally applicable, we must keep in mind that editing is highly contextualized: it always involves specific textual objects that have their own contextual history and takes place in a specific cultural context, and most importantly, for a *specific purpose*. In addition to explicitly acknowledging that this purpose in the case of critical editing has been mainly the facilitation of *literary criticism*, this edition also argues that this should not be seen as the exclusive—or even the predominant—purpose of editing historical texts, and that traditional critical editions must thus be seen as only “one among a range of options” (Hanna 1992: 129). Similarly, the reduction of a *work* into a “a single textual product” should not be seen as “an unarguable given” (Eggert 1991: 65) or an end in itself, but rather as a practical tool for a specific purpose, which in the case of critical editing has been the literary analysis of that work.

From the point of view of the present edition, the methods of critical editing are seen to be unsuitable in three respects. First of all, their assumptions about an authoritative original text and its transmission do not reflect the reality of practical medieval texts transmitted as mutable and adaptable discourse colonies. Second, their focus on the abstract *work* instead of its textual and documentary realizations places their interests at odds with those of linguistic study. Third, many of their principles and practices are based on and dictated by the limitations and requirements of the printed medium and are unnecessary—and often counterproductive—in a digital editorial context. These problems are of a more fundamental nature than those outlined by Shillingsburg (1991: 23) within the paradigm of critical editing, i.e. the questionable validity of critical editions as historical constructs, the variety of competing editorial orientations, and the question whether the work should be represented as a product or a process, because they result from the fundamental principles of critical editing and cannot thus be solved while holding on to this editorial paradigm.

### 3.1.2 Critical editing of Middle English utilitarian manuscripts

In the generalized authenticity of the medieval work, all that philology could see was lost authenticity. Medieval philology is the mourning for a text, the patient labour of this mourning. It is the quest for an anterior perfection that is always bygone, that unique moment in

which the presumed voice of the author was linked to the hand of the first scribe, dictating the authentic, first, and original version, which will disintegrate in the hands of all the numerous, careless individuals copying a literature in the vernacular. (Cerquiglini 1999: 34)

The origins of modern textual criticism in Renaissance ideology and the disciplines of classical and biblical studies means that its application to Middle English texts, produced in a very different and “at times, even overtly antithetical” context is, to put it mildly, problematic (Machan 1994: 18). This has not prevented textual criticism from becoming the dominant “discursive field” of Middle English editing (Machan 1994: 178), resulting in the implicit acceptance of the rather dubious assumption “that in the Middle Ages literature existed textually, aesthetically, and culturally much as it does in the modern (or Antique) period and that the goal of the editor of medieval works is consequently much the same as the goal of someone concerned with other literature” (Machan 1994: 60).<sup>11</sup> Since the traditional methods of textual criticism were originally developed for reconstructing the archetypal text of biblical and classical texts in Latin, and subsequently adapted for recovering the authorial text of modern *literary* works for the purposes of literary criticism, we should not be surprised to find them ill-equipped for dealing with medieval non-literary and utilitarian texts. The fact that critical editions of medieval works—even non-literary ones—have been and are still being produced<sup>12</sup> is not only somewhat surprising but also indicative of two things: editors rarely scrutinize the assumptions underlying established editorial practices, and no editorial approach has been developed specifically for the editing of utilitarian texts surviving in manuscript form.

### Authorial originality and scribal corruption

While discussion about the mismatch between textual criticism and medieval textuality often focuses on *methodological* issues, the most fundamental problem with the critical editing of medieval textual objects, especially utilitarian ones, is not methodological but *ontological*.<sup>13</sup> This fundamental problem derives from the fact that—as described in section 2.3—“Middle English scribal transmission is not, in fact, congruent with traditional textual criticism’s conception of variation” (Machan 1994: 168). Whereas the fundamental principles of critical editing are

<sup>11</sup> Machan’s formulation of these assumptions is interesting, because it reveals a further implicit assumption, namely that editors of medieval texts are interested exclusively in *literary* texts, leaving editors of medieval non-literary texts—like the present editor—doubly marginalized.

<sup>12</sup> For example the Early English Text Society’s current guidelines (EETS Guidelines) state that the society “currently prefers critical editions, offering a single text, into which emendations have been incorporated, and not relegated to the critical apparatus” (3), and editions of Middle English literature in the Norton Critical Editions are essentially copy-text editions.

<sup>13</sup> The practical and methodological problems involved in editing medieval manuscripts have been discussed for example by Pearsall (1985), Allen (1987), Shillingsburg (1991), Brewer (1992), Pearsall (1994), Edwards (1998a), Moffat (1998), Cerquiglini (1999), and Williams and Abbott (1999), and more specific problems related to the critical editing of utilitarian texts and other texts exhibiting a weak author function and extensive textual variation by Edwards and Moffat (1998), Fellows (1998), and Hieatt (2004), among others.

based on an assumption of a strictly monogenous tree model of textual transmission, medieval textual transmission would seem to have been anything but. The central assumptions common to all approaches to critical editing that are the most relevant in this respect are:

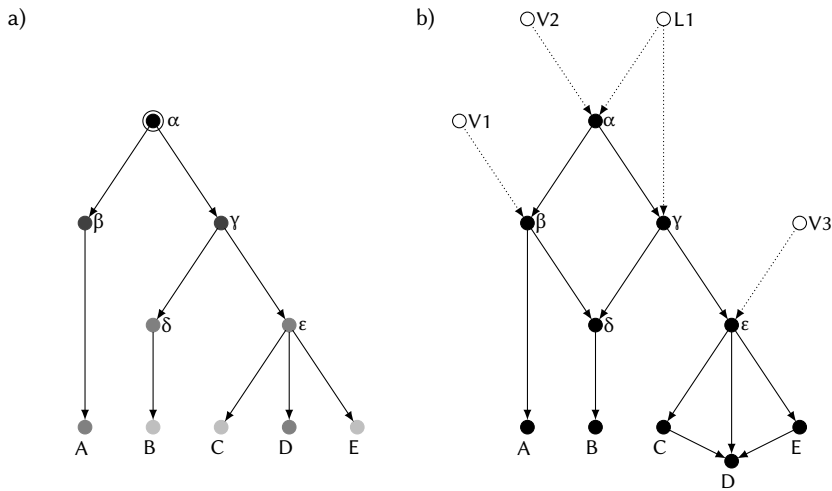
- 1) that a *work* is an autonomous entity with a distinct identity, created by an original act of composition resulting in a finished ‘authorial text’, the textual realization of a unique *authorial intention*;
- 2) that the scribes copying the work base their work on a single exemplar and aim at its faithful reproduction, including its accumulated errors, but always introduce a varying amount of new errors, resulting in a continuous degeneration of the text;

These two basic assumptions result in a ‘tree-like’ conceptual model of textual transmission that looks like *Figure 3.1a*, branching out from a single ‘root’, the *authorial text*.<sup>14</sup> Since critical editors see this authorial text as an ‘original’ creation, the purest representation of what is seen as the essence of the *work*, and thus also qualitatively ‘better’—both in linguistic and literary terms—than any scribal reproduction of it (Embree and Urquhart 1987: 52; Allen 1987: 21), it becomes the natural object of interest for the editorial project. The task of the critical editor—a quite reasonable one in the light of the above assumptions—is then either to recreate a version approximating  $\alpha$  as closely as possible, or if such a reconstruction is deemed impossible (the ‘best-text’ position), to present a version of the best surviving witness (which would here be *A* or possibly *D*), purged of obvious errors. However, the problem with medieval manuscript texts, especially ones which have a utilitarian function, is that neither of the assumptions outlined above would seem to hold.

The first of the above assumptions is intimately tied with the conception of an *author* as an individual and autonomous artist, a literary and linguistic genius, whose *intention* in producing the work is seen as the progenitor of the whole textual tradition of the work. As was pointed out in section 2.3, this text-critical figure of the author has very little to do with medieval vernacular text production, which was less a process of original creation than one of reformulating pre-existing knowledge, whether received from earlier writings (in their own or a foreign language), from oral tradition or from common experience: few if any Middle English texts “were ‘original’ in the sense ‘not derived from something else’”, the expectation being “that any Middle English work *would* be derivative to a greater or lesser degree” (Hamel 1998: 204).<sup>15</sup> This means that at least in the context of non-literary works, the medieval *author* should be seen less of an ‘originator’ than a ‘mediator’, and the *work* should in turn be seen less as an independent artistic

<sup>14</sup> In the figure, the saturation of the dots represents the amount of corruption—in terms of accumulated copying errors—in relation to the authorial original ( $\alpha$ ), and thus their relative *authority*.

<sup>15</sup> For example in the context of medieval medical treatises, Pahta (1998) has observed that many of them are “not original compositions in the modern sense of the word, but conflated texts assembled from various sources” (16). Similarly, Grund (2006) has observed that Middle English alchemical texts often vary widely, “most likely since they were produced by knowledgeable copyists and practitioners who revised their exemplars in accordance with their own experience or reading of other sources” (106), sometimes replacing entire recipes with new ones copied from somewhere else, “perhaps because he felt that the recipe in his exemplar was incorrect and/or because he had a better recipe at his disposal” (108).



**Figure 3.1:** *Models representing textual transmission as assumed by textual criticism (a), and reflecting current understanding of medieval scribal practices based on the study of surviving medieval manuscripts (b).*

creation than a constantly evolving conglomeration of socially produced knowledge.<sup>16</sup>

Along with the valorization of the *author*, critical editing also tends to valorize the *editor*, portraying his work as “an extended demonstration of literary judgment, involving a sublime notion of the author as a unique someone whose words will always be discerned to be ineffably superior among the mass of scribal variants, and a sublime notion of the editor as another unique someone who will always be able to discern them” (Pearsall 1992: 43). This “boldness” of the critical method, basing the edited text on “the opinions of an editor living five hundred years after the composition of a work” (Allen 1987: 13) instead of contemporary or near-contemporary documentary evidence reflects the same romantic idea of the exceptional genius as its obsession with the authorial text.<sup>17</sup>

### Unauthorized textuality

The derivative nature of even ‘original’ medieval text-production, together with the fact that medieval scribes frequently translated the linguistic forms of their exemplar to those of their own idiolect,<sup>18</sup> corrected mistakes in their exemplars

<sup>16</sup> This is even more so with texts originating in an oral tradition—such as recipes, which could very well be passed on in oral form long before they were written down—since the first written record would most likely have already been heavily edited in the act of transcription—is already a scribal one by text-critical standards (West 1998: 94).

<sup>17</sup> This same phenomenon is apparent in the rhetoric that bases the perceived superiority of critical editions on the philological prowess they demand of the editor instead of their usefulness for research; diplomatic or other noncritical editions are criticized not so much for being useless, but rather for being too *easy* to prepare, as if the main purpose of editing texts was to show off the editor’s skills.

<sup>18</sup> This tendency has been well documented in linguistic studies of Middle English manuscripts and their transmission histories (see e.g. LALME: 1:13–23, McIntosh 1989b: 92, Laing 2000: 99–100,

(Machan 1994: 171), and modified their contents by adding or deleting material (Keiser 1998c: 112; Voigts 1989: 350), means that “each act of copying was to a large extent an act of recomposition”, “performed at a level of intellectual and imaginative engagement not inferior to and little different from the putative original act of composition” (Pearsall 1985: 101). These kinds of scribal strategies seem to have been especially common with utilitarian texts like recipes, household advice and almanacs, as well as popular literature (e.g. verse romances) that lack the sacrosanct nature of religious works (Keiser 1998c: 112). While some canonical works of literature or philosophical writing—in Latin, naturally—achieved a status resembling that of religious texts, vernacular texts were by definition seen as ‘nonauthoritative’ and thus freely alterable. Thus, for Middle English texts, the stark contrast between the original and creative *author* and the “careless and meddling” (Blake 1998: 64) *scribe* producing merely “pedestrian alterations” (Moffat 1998: 37) is simply not borne out by the manuscript and cultural evidence (Machan 1992: 6).<sup>19</sup>

In discussing a 14<sup>th</sup>-century inquisitorial treatise of Zanchino Ugolini, Diehl (2004) has called these kinds of medieval vernacular texts, “not fixed immutably by tradition” but rather freely added to and changed by their users, “*textes vivants*” (68). This concept a *living text* is well-suited to describing the textual transmission of medieval recipes and other utilitarian texts which do not have a strong author function but are rather based on the accumulated knowledge of a discourse community, being “what we might call performance texts, the residue of social occasions, in which unknown numbers of performers and redactors have participated [...] as equal participants in a continuing creative process” (Pearsall 1994: 125). For these “more pedestrian kinds of composition [...] it becomes clear that the assumptions on which the critical edition is based—concerning the coherent and unified process of composition, the definitive textual moment, and the act of publication and authorization—are not merely difficult to apply but irrelevant” (100).

Furthermore, vernacular texts, especially utilitarian ones like recipes, were often copied for their own use by amateur scribes—i.e. literate laymen—who copied texts for their own pleasure and use and were not significantly influenced by the principles and practices of copying established for religious and literary texts by professional and monastic scribes, further widening the range of variation found in these kinds of texts (Allen 1987: 15-8; Fahy 2004: 402). Thus, as Pearsall (1985: 102) has demonstrated in the case of a prose translation of the pseudo-Bonaventuran *Meditationes Vitae Christi*,<sup>20</sup> it seems obvious that there is no reason to see the authorial version of most medieval texts as inherently ‘better’ in terms of its language or content. For this reason, this edition assumes the social-textual position advocated by McGann (1991 and 1992) and Cerquiglini (1999), which sees textual objects as the products of not individual authors but of specific cultural

---

Grund 2006: 112-7, and Thaisen 2008: 43-7), and has been acknowledged even by many editors and textual scholars (e.g. Blake 1998: 66 and Lucas 1998: 172).

<sup>19</sup> While many critical editors and textual theorists acknowledge the fact that in addition to making unconscious errors, medieval scribes also made conscious changes in copying, they still consider these changes to be ‘errors’ (e.g. Kane’s concept of “conscious errors”).

<sup>20</sup> Of the surviving manuscripts, one subgroup is clearly superior in terms of syntax, vocabulary and sense, yet “it is clear that the inferior subgroup, crabbed and awkward in style, defective in syntax, often unintelligible, is the original” (Pearsall 1985: 102).

and situational *contexts*, as defined in section 2.2, of which the author is merely one component.

Thus, instead of the sharp distinction between the originating author and the corrupting copyist implied by the tree structure of *Figure 3.1a*, we have a network of textual ‘consumer-producers’ who make use of existing textual resources to produce new texts bearing varying relationships to their predecessors. Some of these texts can be considered to represent *versions* of the same *work*, while others are perceived as entirely new derivative works. This kind of a *rhizomatic* (Deleuze and Guattari 1987) view, which is here argued to be a more accurate description of medieval textual transmission, is represented in *Figure 3.1b*. In light of this kind of a model, where the textual history of an individual *work* has no definite beginning or end, but is instead a part of a larger textual network, the concepts of *error* and *corruption* become irrelevant since there is no ‘authorized’ point of origin to serve as a reference and are replaced by the more neutral concepts of *variant* and *variation*.<sup>21</sup> Additionally, this model also differs from that assumed by textual criticism in acknowledging the fact that medieval scribes could well make use of more than one exemplar in copying a work (Hanna 1992: 120), producing ‘edited’ textual versions that combine material from several exemplars (as in the case of  $\delta$  in the example) or emend a principal source text (in essence serving as a *copy-text*) on the basis of other available manuscripts (*D*, based on  $\varepsilon$  emended on the basis of *C* and *E*).<sup>22</sup> In addition to this work-internal lateral transmission (use of multiple exemplars), the rhizomatic nature of medieval textuality is also reflected by *Figure 3.1b* in the form of the empty circles (*V1*, *V2*, *V3* and *L1*) representing other works that have provided material for the different textual versions of the work represented by the diagram.<sup>23</sup>

It is precisely this lack of substantial distinction between original *authors* and mere *scribes* that sets these kinds of ‘unauthorized’ works apart from modern printed literary works on which much of text-critical theory is based. Because their transmissional history does not contain a determinate moment of publication, they lack the kind of a definite identity established by the printing industry during the early modern period and implicitly assumed by nineteenth-century philologists (Cerquiglini 1999: 34). Instead, they essentially remain in what could be considered either a perpetual pre-publication state<sup>24</sup> or a state of constant re-

<sup>21</sup> In the figure, this is represented by the absence of colour gradation, each textual version being judged on its own terms as a reflection of the particular *textual*, *situational* and *cultural* contexts.

<sup>22</sup> This horizontal dissemination of material resulting from the use of multiple exemplars is called ‘lateral contamination’ by textual theorists, and is seen to violate the text-critical model of textual transmission, which “presupposes a strictly vertical descent of manuscript copies” (Moffat 1998: 30). Consequently, the occurrence of this kind of horizontal dissemination, “for which there is considerable evidence - especially in the more popular works” (Moffat 1998: 30), is seen by many scholars to invalidate the use of critical editorial approaches for such texts.

<sup>23</sup> In this hypothetical example case,  $\alpha$  is a translation of a Latin work (*L1*) that incorporates additional material from an earlier vernacular work (*V2*), forming what is considered to be the earliest version of a new work.  $\beta$  and  $\gamma$ , then, are derivative copies of  $\alpha$ , the former incorporating further material from another vernacular text on the same topic (*V1*) and the latter being a partial retranslation of *L1* which nevertheless copies the majority of its material from  $\alpha$ .

<sup>24</sup> As Pierazzo (2009)—following Dirk Van Hulle—has pointed out, this should not be taken as an excuse to equate medieval manuscripts with modern unpublished or pre-publication manuscript texts, as they are crucially different in that while medieval manuscripts were a vehicle of communication, intended to be publicly disseminated, the introduction of printing gradually relegated manuscript texts to the private domain. Thus we should be wary of applying to medieval texts editorial theories—like

publication, perpetually mutable (at least until fixed by a modern ‘authoritative’ edition) both in their form and content and available for free appropriation and reworking by each successive scribe-author. In the case of recipe collections like the *Potage Dyvers*, this issue of indeterminate identity is compounded by the structurally fluid nature of the text as a *discourse colony*, which allows and even encourages individual recipes to be left out, added, replaced by superior parallel versions, or independently modified. This forces the editor to account for variation on two separate levels: not only are there variant versions of individual recipes, but different versions of the collection as a whole also contain variation in their selection and ordering of those recipes (see chapter 13). Thus judgements about the identity or non-identity of individual recipes based on the degree of their internal variation also escalate onto the macro-level of the entire collection, affecting judgements on the degree of identity of the variant versions of the entire collection. This radically variant and non-self-identical nature of the medieval work is something that critical editing, oriented towards the “completed version, ready for the press, authenticated and authorized” (33), is simply not equipped to deal with:

When this practice turns to medieval manuscripts in the vernacular, the philological automaton gets carried away and panics. There is so much variability in number, extent, and nature of the readings that the work is immense and success illusory; the whole, vast undertaking seems maddening and humiliating. (Cerquiglini 1999: 37)

Thus the traditional approach of critical editing, the establishment of a single ‘authoritative’ text representing the intention of a single author—either through the reconstruction of an archetype or the preservation of a single ‘best text’—is an impossible—and irrelevant—goal in the case of these ‘live’ and ‘unauthorized’ texts. Instead of trying to eliminate this variation as mere ‘scribal corruption’, it would be more fruitful to see it as a result of ‘scribal adaptation’ and *reauthoring*—in which misreadings and misunderstandings naturally play a part—and to produce editions that highlight this variation and help us learn more about medieval textual practices and about “how the scribes understood fidelity and whether they saw it as their primary objectives” (Machan 1994: 169). As “critical edition is by its nature resistant to and incapable of coping with such texts”, Pearsall (1994: 111-2) sees the only practical solution to be a parallel text edition of the surviving witnesses, which is also the basic approach taken by the present edition and will be discussed in more detail in subsection 3.2.3. From this point of view, shared by the present author with McGann (1985) and Pearsall (1994), *textual criticism* should thus properly be seen not as an editorial methodology, but a literary critical one. As a historically grounded discipline, separate from and posterior to editing, its purpose is to create “an understanding of the entire developing process of a literary work’s historical transmission” (McGann 1985: 191-2), much in the same way as dialectal analysis is used to gain information about the linguistic and geographical background of manuscript documents. What thus remains as the task of scholarly editing is the provision of useful and analytically powerful representations of historical textual objects for the purposes of textual criticism and other kinds of literary, linguistic and historical research.

---

genetic editing—that have been developed for modern manuscript texts in mind.

### 3.1.3 Implications of critical editing for linguistic research

Not only as a linguist, but as someone who enjoys reading mediaeval literature, I find myself at a loss to account for much of what textual editors not only do, but admit to doing. Perhaps this is a failure of willing suspension of disbelief on my part, but I find the characteristic description of how one ‘arrives at a text’ (even the idea that such an activity should count as legitimate scholarship, however elaborate and learned it is) the sort of thing I would not like to be known to be doing. (Lass 2004b: 37)

There has been much discussion about the suitability—or rather unsuitability—of critical editions for linguistic research over the last decade or so, perhaps the most widely known being the rather scathing attack of Lass (2004b), quoted above. Since the present edition is conceived specifically as a linguistically oriented edition, this section will examine the problems inherent in the relationship of critical editing and historical linguistics. The problems caused by using critical editions as data in linguistics, first pointed out in the 1980s, are beginning to be commonly acknowledged at least by historical corpus linguists and compilers of linguistic text corpora, although practical concerns have until very recently forced historical linguists to rely on printed critical editions perhaps more than they should.<sup>25</sup> Although the compilation of *historical corpora* (see chapter 4) over the last two decades has made it more obvious, the linguistically problematic nature of reconstructive critical editions is hardly a recent discovery. Already in 1872, when Henry Nicol wrote an otherwise favourable review of Gaston Paris’ edition of the *Vie de saint Alexis* he “questioned the value of critical texts for works such as translations, whose interest is mainly linguistic” (Foulet and Speer 1979: 16), already recognizing that from a linguistic point of view, the practice of critical editing produces “fictitious objects representing nothing of historical interest” (Lass 2004b: 24).

As was already mentioned above, the very core concern and principal justification of critical editing—the desire for the authorial text—is a characteristically literary concern. Whereas it is understandable that a literary scholar interested in the *oeuvre* of a specific author wants to find out what that author him- or herself wrote, historical linguists are not generally any more interested in the language use of an ‘author’ than in the language use of a scribe.<sup>26</sup> They *are*, of course, interested in identifying and isolating the language use of one individual from that of others, but from a linguistic point of view, the spontaneous language use of a scribe is just as interesting as that of the ‘original author’. This results in a rather

<sup>25</sup> The reason for this, of course, is that as for example Carlquist (2004: 112) and Grund (2006: 110) point out, the majority of Middle English editions available even today are literarily oriented critical editions that focus on the abstract *work* and its reconstruction instead of presenting a manuscript version of the work ‘as is’.

<sup>26</sup> The only reason for such a specific interest—especially for sociolinguists—would be the fact that there is often more socio-historical information available for people who were prominent in their own time—which includes some authors.



drastic change of perspective: where a literary scholar sees a document as containing the *work* of the author, *corrupted* by the interventions of a series of scribes reaching down to the current scribe, a historical linguist sees it as containing a *text* written by the scribe influenced by a series of preceding copies, reaching back to the authorial copy. This means that the authentic surviving linguistic usage of the scribe should always be preferable to the hypothetical reconstructed linguistic usage of the author.

### Editorial language

The most obvious problem of critical editions from a linguistic point of view is the tendency of editors to interfere with the linguistic form and content of the text. The most blatant example of this is the second stage of the Lachmannian recensionist method, following the establishment of the most authoritative readings, which involves the editor rewriting the composite text “to make it conform to the author’s own language—where that may be determined—in morphology, phonology, syntax, and dialect, which no manuscript preserves perfectly” (Foulet and Speer 1979: 10), as this naturally makes the edition completely worthless as linguistic *evidence* and more representative of a specific linguistic hypothesis. Although this kind of full Lachmannian reconstruction is fortunately rare, the commonly held view in all types of critical editing that “[a]ny manuscript one chooses as base will need grooming, what the French call ‘la toilette du texte’” (Foulet and Speer 1979: 76) conflicts with the linguistic view that in order to be “authentic”, the text should be “as faithful a transcription as possible, ‘warts and all,’” (Moffat 1998: 33), making the critical edition unable to serve as linguistic evidence. While this ‘grooming’ is often defended on the grounds that all emendations made to the text are indicated by the editor, it is generally acknowledged that it is not possible to indicate all the editorial interventions in the reconstructed text.<sup>27</sup> Furthermore, many critical editions go even further than this and present what is known as a *clear text* containing “neither critical symbols or footnote numbers to indicate that an emendation has been made or that some detail has been omitted” (Kline and Holbrook Perdue 2008: 173).<sup>28</sup>

The problem here seems to lie in the different conceptions of *text* in the literarily oriented field of critical editing and linguistics. Whereas linguists, in terms of the classification presented in section 2.1, are naturally interested in the linguistic *text* as an encoded record of a language event, literary and thus also textual critics are traditionally interested in the abstract *work* underlying the surviving documentary forms. This means that even relatively conservative editors have

<sup>27</sup> The common practice in the critical editing of literary works is thus to *silently emend* many aspects of the text—like capitalization and punctuation, as well as orthographical details (Kline and Holbrook Perdue 2008: 145). These kinds of editorially inserted features are not indicated in the text but simply noted in a statement of editorial principles preceding the text, for the practical reason that it would be “unduly burdensome to indicate all such punctuation within square brackets” (Edwards and Moffat 1998: 230).

<sup>28</sup> The virtue of *clear text* presentation is usually seen to be “that it offers the readers no distraction in their perusal of the text” (Edwards and Moffat 1998: 235), but it has also—quite rightfully—been criticized for privileging the editorially constructed text over the texts of the surviving documents and thus alienating the reader from the material nature of the original documents (Hanna 1992: 111), and for presenting the editorial text as the “original document of which the actual medieval texts are the degenerate copies” (McGillivray 1994: 180-1).

been perfectly comfortable with normalizing those aspects of the text that they consider nonsubstantive in terms of the meaning of the *work*:

While many editors today respect the faithful reproduction of the historical form of texts as a primary requirement of critical editing, they act differently for those peculiarities of manuscript or typescript witness documents that reflect the respective situation of the writer. Such specifics are often normalized, although they may refer to a pragmatic context of the genesis of the text that can have far-reaching consequences for an adequate evaluation and interpretation at a given textual state. (Martens 1995: 218-219)

This tendency is tied closely to the perception of critical editing as the establishment of a single, stylistically uniform text, emended to reflect as closely as possible the *work* the author intended. The problem is that this aim is diametrically opposed to the *variationist* view of language typical of historical linguists (see chapter 4), which is interested precisely in the variation between different linguistic realizations of the same *work* produced at different times and in different situational and cultural contexts. This interest in *contextualized* language use makes critical editions even more problematic for pragmatically oriented historical linguistics, because in attempting to reconstruct the abstract *work*, a critical edition radically decontextualizes the *text*. By abstracting the original *documents* into a *work*—or a *version* in a more conservative edition—and then reconstructing for it new textual and documentary realizations conforming to modern printed conventions, a critical edition obscures—or worse yet, falsifies—whatever contextual and paratextual information the original documents contained:

For students of manuscripts the critical edition, as has often been pointed out in recent years, leads only to radical decontextualization, removing the text itself from any connection with the forms in which it was transmitted, whereby texts are isolated from crucial facts of their production that have potential bearing on their meaning: what forms of script were used to copy it; how many scribes; what materials were used in its construction; if (and if so how) it was decorated; it [sic] what dialect(s) it may have been written; and whether it was collocated with other texts and whether they are verse or prose. In this sense what we have is clearly an act of dehistoricization, one that poses some questions about our modern relationship to medieval manuscript artifacts. (Edwards 2000: 75)

### Textual eclecticism

While the problem of reconstructed linguistic forms may be the most obvious and destructive flaw of critical editions in terms of their linguistic witnesshood, many scholars have argued that a more insidious and therefore more serious problem is constituted by the hybrid nature of all reconstructive critical editions.<sup>29</sup> As Grund (2006) points out, even without normalized linguistic forms or conjectural

<sup>29</sup> Best-text editions can avoid this problem as far as they resist emending 'errors' in the best text on the basis of other witnesses.

emendations, the procedures of critical editing produce linguistically composite texts that have lost the authenticity of the scribal versions, and prevent the user's access to real historical utterances and by presenting a modern sampling of an entire textual tradition which has never existed "all together in the same place at one time" Lass (2004b: 37).

This conflation occurs along both the chronological and spatial dimensions. First of all, the critical editor who includes in the text readings from different manuscript versions "contaminatingly synchronizes that which occurred diachronically", depriving the work "of its historical dimension in projecting all distinct versions onto one level and producing out of them a single, new version in the name of authorial intention" (Zeller 1995b: 106).<sup>30</sup> Second, even if the editor includes readings only from manuscripts roughly contemporary with each other, he or she is still conflating utterances produced in potentially very different circumstances by persons of different sociocultural and linguistic backgrounds, producing a sociolinguistically and dialectally problematic text.

The traditional device used by critical editors to defend against these allegations of hiding the documentary evidence is the *editorial apparatus*, which contains those substantial variants *not* selected for the text itself. This "apparatus of allographic notes" (Genette 1997: 337) "is supposed to reproduce all authorized states of the text" (Zeller 1995a: 27), preserving a record of the range of textual variance. However, in practice the range of variation preserved in the apparatuses of critical editions has been far from complete, allowing the reader access to "a few fragments, splinters and scraps, but not the other of the text" (Cerquiglini 1999: 72-3).<sup>31</sup> Many scholars have also questioned the function of the critical apparatus as a source of information about the text, arguing that its real function is in fact "not to inform the reader but to protect the editor" (Vanhoutte 2009: 106).<sup>32</sup>

<sup>30</sup> This means not only that the edited text may contain linguistic usages from radically different time periods (especially medieval texts can have textual traditions centuries long), but that it may contain usages from different 'evolutionary versions' of a text, juxtaposing utterances that at worst reflect entirely different communicative strategies.

<sup>31</sup> There are two main reasons for this failure. First of all, conventional critical editions exclude all variants that are considered mere spelling variants from the apparatus, "because orthographic variation is unlikely to weigh heavily in the establishment of the text" (Edwards and Moffat 1998: 229). Second, the traditional critical apparatus is "designed for transmission of variants in versions that are characterized by more identity than variation" (Van Hulle 2006: 158) and "does not normally allow for anything other than the barest reporting of manuscript evidence" (Edwards and Moffat 1998: 232), simply failing in the face of structural or extensive lexical variation of the kind found in medieval utilitarian manuscript traditions.

<sup>32</sup> This editorial focus of the apparatus has also been observed by McGann (1991) and Lavagnino (2009), who point out that the editorial apparatus has over the last 50 years focused on the activity of the editor rather than on the textual history of the edited work. This should not come as a surprise, for as Vanhoutte (2009: 105-6) reminds us, the preservation of documentary evidence was never its historical purpose: for example Bowers (1959) "did not consider it a function of the critical edition to reproduce the textual evidence 'since the editor should have exhausted their significance in the preparation of the definitive text'". As Vanhoutte observed, this level of confidence in the editor's critical abilities flies in the face of the scientific principle of external replication, as the apparatus of a traditional critical edition is the only source of data for the assessment of the validity of the editor's decisions.

### Fallacy of textual reconstruction

Although the problems mentioned above can be seen merely as results of the established *practices* of critical editing—which have historically been much less established than the underlying theory—many critics of textual criticism have pointed out that the very goal of critical editing, namely the attempt to reconstruct the ‘archetypal’ textual state, is fundamentally incompatible with the linguistic view of a *text*. The reason for this is—as both Machan (1994) and Lass (2004b) have argued—the failure to properly distinguish between the different ontological levels of the text, described in section 2.1. This has led to the equation of the *work* with “the correct text” (Machan 1994: 138), and to it being seen as *ontologically similar*—even if substantially different—to the texts manifested in the surviving documents, instead of being properly recognized as a higher-level abstraction of them. This ontological misunderstanding is basically what lies at the root of the misguided attempt to reconstruct the original *textual realization* of the work: because the *work* exists purely on the ideational level, it *cannot* be used as a guide for reconstructing a *linguistic* realization for it.<sup>33</sup>

This misguided endeavour is what Lass (2004b) has rather scathingly labelled as a “lethal epistemological disease” and “a venerable reconstructive error” (29). While the text-critical project of recovering an archetypal version of a text was inspired by the project in 19<sup>th</sup>-century comparative linguists of reconstructing the archetypal forms of languages, the analogy “is fallacious for an elementary structural reason” (Lass 2004b: 29). While the “reconstruction of lexis, syntax, morphology, phonology is done at the (rather abstract) ‘system’ level”, the “reconstruction of particular texts is of necessity an attempt at *the reconstruction of utterances*”, which is “not attainable with any epistemological safety” (29). The reason for this is the fact that “utterances (prose works, poems, letters, shopping lists, marginalia, graffiti) are contingent and spatiotemporally located objects”, not merely the products of a *linguistic system*, but “of *a particular contingent deployment of this system by an individual human being*” at a particular place and time (Lass 2004b: 29), as described in chapter 2. Thus even if it is theoretically possible to reconstruct the ‘original’ or ‘archetypal’ ideational *work* on the level of semantic content, it is epistemically impossible to reconstruct the specific linguistic and orthographic forms in which that archetypal *work* was expressed.

Lass (2004b) has argued this problem is based on a “*confusion between cladogenesis and reconstruction*” (30), or the assumption that plotting the stemmatic relations of a set of manuscripts would allow one to say something about the shape of their archetype. However, since the stemma is a “purely relational phylogenetic tree or genealogy” which does not “contain any information about details of the original that fail to be inherited” (30), any reconstructed *textual* archetype is fundamentally flawed.<sup>34</sup> Despite his polemical tone, Lass emphatically points out

<sup>33</sup> Although it can be used as a criterion for discriminating between different surviving linguistic realizations in terms of their correspondence with the ideational *work*.

<sup>34</sup> In order for the linguistic reconstruction of an earlier archetype to be possible, the exact linguistic form of every single word in the archetype would need to have been preserved unchanged in at least one surviving manuscript, *and* the editor would need to be able to somehow differentiate this original form from all the other forms into which the scribes of the other surviving witnesses translated it. In the light of the preceding discussions about medieval scribal practice, this is astronomically unlikely.

that he is “*not* deprecating reconstruction or extrapolation as necessary weapons in the historical linguist’s armoury”, but objects “to the reconstruction of *texts*, and their subsequent use as witnesses for linguistic study – not to reconstruction as a source of knowledge of earlier language states at ‘system’ level” (24). Thus, in terms of the diagrams presented in chapter 2 (*Figures 2.2 and 2.3*), it is theoretically possible to reconstruct objects higher up on the scale of abstraction and instantiation based on those lower down (i.e. to move upwards and leftwards in *Figure 2.2* and upwards in *Figure 2.3*), but not the other way around.

The severity of these limitations naturally raises the question of why are critical editions used for linguistic study in the first place? The answer is as simple as it is depressing: availability. Because of the historical prevalence of the Greg-Bowers school of editing, there are simply many more critical editions than editions of any other kind. Since few historical linguists or corpus compilers have the resources to re-edit texts from the original manuscript sources using methods optimally suited for their research, they compromise and make do with what is readily available in sufficient quantities: critical editions. In order to help remedy the situation, chapters 4 and 5 focus on ways of creating editions that break with those aspects of traditional editorial wisdom that are harmful for linguistic research without unduly sacrificing the usability of the edition for other purposes.

### 3.1.4 Critical editing and the digital medium

It is a commonplace that modern critical editions of medieval works, with their orthographic conventions, elimination of copying errors, reader aids, and uniformity, differ radically from medieval works, which materially exist in physically and lexically variable manuscripts.

(Machan 1994: 65)

The discipline of critical editing is inextricably tied with the modern printed book with its “particular script and print-based technologies and distribution logistics”, based on its specific “media materialities and epistemologies” (Dahlström 2009: 27). Consequently, its methodology is to a large extent a means of transforming historical works, surviving in the form of a varying number of *textual versions* encoded in physical *documents* of varying characteristics, into a printed book that fulfils the expectations of the modern reader and allows her or him convenient access to what is seen to be the textual essence of the historical work. This means that in representing the historical work, critical editors have had to develop their methods not only in accordance with their ontological views of what the text is and the uses they have envisioned for the edition, but also in accordance to the physical and cultural limitations of their *target medium*. Since both modern textual ontologies and the ways in which scholars use editions are to a large extent dictated by the physical form of the printed book, the characteristics (and limitations) of the printed book are in fact responsible for many of the practices of traditional critical editing. For this reason, the recent move of scholarly editing into the digital domain with its radically different characteristics is in itself a reason for critical re-evaluation of the established conventions of scholarly editing.

### From manuscript multiplicity to printed unity

As for example Greetham (1991) and Eggert (1994) have argued, the reduction of a multitude of manuscript texts into a single critical reading text—the definitive function of critical editing—is a practice dictated as much by the “pragmatic and commercial constraints” of print publishing “as by intellectual commitment” (Eggert 1994: 6).<sup>35</sup> However, as the advances in digital representation technologies have to a large degree obliterated these constraints, the case for critical editing, frequently justified by practical arguments, has been significantly weakened.<sup>36</sup> The coincidence of the gradual recognition of the digital medium as a viable—and over the last decade, even preferable—medium of scholarly editing with the acknowledgement of the multifarious nature of (especially medieval) texts has led many textual scholars to abandon the editorial quest for a singular text and to concede “that critics—and earlier editors—had their entire object of study wrong: they thought a work was a single thing, when it’s really a whole range of texts and their relationships” (Lavagnino 1995: 112).

As long as printed editions were the only kinds possible, the relationship between editorial theory and its material medium was rarely discussed, but the radically different logistics and parameters of document production in the digital medium have forced editors and textual theorists to pay more attention to it, resulting in more and more scholars no longer finding the traditional critical edition the only acceptable—or even the most suitable—goal for an editorial project (Ore 2004: 36, see e.g. Renear 2001 and Vanhoutte 2006).

### Textual abstraction

Another significant result of the printed book as the default medium of critical editions has been the reduction of medieval textual objects into their abstracted textual content, which is the only aspect of them that can be conveniently represented within the conventions of the modern printed book. As Hanna (1992: 111) points out, most critical editions pay no attention to the physical attributes of the source manuscripts—often reduced to mere sigla—disregarding the fact that the contents, material, script, dialect and various paratextual features of manuscripts can significantly affect our evaluation of them.<sup>37</sup> By separating the encoding of

<sup>35</sup> Johansson (2004) has defined the practical limitations imposed by the printed book in terms of *space*, *selection*, and *time*, the common factor being *economy*. In the ideal world, a printed edition would “use as much space as is needed to print all text witnesses, organize the material in various ways, and use the time needed to do this” (94), but since this is usually not economically viable, editors have developed various ways of reducing the multiplicity of documentary witnesses into a single linear text conforming to the modern ontology of the printed text.

<sup>36</sup> Howard-Hill (1991) presents an example of the practical justification of critical methods. He argues for the necessity of making the traditional editorial choices because the publisher insists on them, quoting “economic and market forces” as the restrictive factor preventing an editor from providing multiple versions of texts or adopting “complex or novel arrangements of textual material” (53). While these kinds of purely practical and ephemeral considerations most likely represented the prevailing situation in 1991, using them to justify *prescriptive* rules for preparing an edition seems rather strange. Surely these kinds of practical obstacles should rather provoke attempts to overcome them than enforce an acceptance of prevailing practices, which are recognized as hindering the scholarly potential of editions.

<sup>37</sup> It is somewhat ironic that this “concentration on the linguistic text” has resulted in the misrepresentation of the textual object as a material witness to a communicative event by ignoring “the

data from its presentation, the digital medium vastly expands the capability of the edition to store data and allows it to encode multiple layers of data about various aspects of the original document. The encoding of the material paratext on the same ontological level with the textual content allows not only the selective and interactive visualization of various aspects of the textual object, but also the filtering, searching and organizing of the encoded data.<sup>38</sup>

The critical edition—and especially its characteristic critical apparatus—is thus fundamentally a method of reducing multiple variant versions of a work into a single text, so that it can be printed and used as a vehicle of literary criticism. The very existence of the *apparatus criticus* as a required component of the scholarly critical edition reveals it as a compromise: even for critical editors, the single established text is not considered to be enough, but must be supported by as much evidence of variation as can be conveniently represented on the printed page. In the digital realm, there is less need for such a compromise: the separation of data from its presentation, together with hyperlinking and the dynamic, visuospatially flexible and polychromatic representation of text made possible by modern digital technology allows the editor to encode and visualize the full range of textual and paratextual variation in all but the most complicated cases. This does not mean that the editor could or should not establish a preferred version of the text using whatever principles he wishes in addition to the representation of all documentary witnesses, but it does mean that there are no longer excuses for obfuscating a large part of the evidence and privileging the editorial interpretation to the exclusion of others.

## 3.2 Documentary editing

What I want to emphasise is that the allocation of effort between the study of critical editing and the study of non-critical editing is in inverse relationship to their relative *practical* importance. To put this rather crudely, and with deliberate provocation, there is a sense in which non-critical editing is extremely important in our culture and society, while critical editing is in fact relatively unimportant. To have a distribution of attention, and, yes, resources, that not only fails to respect that difference, but actually reverses it, is a very bad thing. It is at least not rational or prudent; it is possibly, arguably, unethical.

(Renear 2001: 33)

Although critical editing has constituted the dominant paradigm of textual editing for the last century, it has always been paralleled by the less ‘idealistic’ practice of *noncritical*, *diplomatic* or *documentary* editing. Unfortunately, these forms of editing have so far received relatively little attention and resources from the textual editing community. Vanhoutte (2006) sees the main reason for this neglect to be “the lack of satisfactory ontology of the text on which the methodology

---

transmissive or communicative aspects of linguistic events” (McGann 1991: 57) and the *paratextual* or *bibliographic* codes that participate in the communication process.

<sup>38</sup> This central aspect of the present edition—and digital editions in general—will be discussed in depth in chapters 4 and 5.

of noncritical editing can be modeled” (170). The present edition and the accompanying discussion is an attempt to establish one model for noncritical editing based on the textual ontology described in chapter 2. This section will briefly review the tradition of documentary editing and evaluate the degree to which its practices are suitable for editing Middle English utilitarian texts such as recipes for the use of historical linguistics.

As several scholars (e.g. Pierazzo 2011, Haugen 2004 and Renear (2001)) have pointed out, faithful diplomatic editions have been—and often still are—seen as mere surrogates of facsimile editions instead of serious scholarly products, and documentary editors as something less than ‘real’ editors. For example Duggan (1996) has characterized editors who “simply transcribe the manuscript” as “incompetent” (232) and rather arrogantly declared that “it is the proper task of librarians to conserve documents and of editors to conserve poems” (233). These kinds of belittling comments betray a view of editing as a monolithic occupation with only one legitimate aim and a single correct method, namely, that of critical editing.<sup>39</sup>

Duggan’s scathing dismissal of modern conservative editing and the equally scathing dismissal of his view by Lass (2004b) illustrates the importance of the *skopos* of the intended edition in the selection of editorial method. From the point of view of Duggan, whose focus is on the metrical choices of the original poet, an edition faithful to the scribal language of the surviving manuscript is just as useless as a text emended on metrical grounds is to a historical linguist like Lass, who is interested in the authentic language act produced by the actual person copying the text. Thus the same edition can simultaneously be seen as extremely faithful and severely misrepresentative, depending on whether one takes as a reference point the linguistic act of the scribe or the intention of the original author. In claiming that modern conservative editors are “by temperament and training committed to naive empiricism” and confuse “the document with the poem”, Duggan (1996: 223) fails to see that there is no confusion: those editors are simply more interested in the actual *document* than in what might have been the original authorial version of the poem.

Despite this long-standing attitude, calls have recently been made even within the literarily oriented community of textual scholarship for more attention to be paid to individual manuscripts (Grund 2006: 110).<sup>40</sup> This attitude, while unfashionable for most of the last century, is not a new one but rather a revival from the 19<sup>th</sup> century, when diplomatic editions were preferred by many editors because while “[c]ritical texts ‘have their day and cease to be’, [...] a diplomatic reproduction, once thoroughly done, goes on for ever” (Rhys and Evans 1887: xv). This same argument has also been presented by modern scholars like Ore (2004), who

<sup>39</sup> This idea of the scholarly inferiority of documentary editing is so strong among textual scholars and editors that even ones like Ore (2004), who are championing non-critical editions as a valid editorial option, often take it as granted that a critical edition will at some point be the eventual outcome of an editorial project: “Editions other than critical editions and archives of such non-critical editions are of value in themselves and are natural building stones for a critical edition whether that is digital or printed” (Ore 2004: 39).

<sup>40</sup> Grund mentions Machan (1994: 190-1) as an early example of this, but already in his 1987 article, Turville-Petre concedes that it is, “of course, quite legitimate to reproduce the documentary form of the work as it exists in one of the manuscripts, as long as that aim is clearly stated and the editor explains his reasons for adopting such a course” (Turville-Petre 1987: 143).



argue diplomatic or ‘archival’ editions to be valid scholarly contributions that in fact have more lasting value than a critical edition.

While the roots of *documentary editing* can be seen to extend to antiquity and to the Middle Ages in the form of verbatim copying of manuscripts, its treatment will here be limited to the practice of representing manuscript texts in a different medium, namely in print or digital form. The earliest stages of documentary editing of Middle English manuscripts, before the mid-19<sup>th</sup> century, have been described by McGillivray (1994) as mostly just “publishing”: an editor “would usually choose what he took to be a good manuscript of a work and transcribe it (frequently introducing numerous errors as he did so)” (175), with the aim of making the medieval text more available to his contemporaries. It was in the mid-19<sup>th</sup> century that editors began to pay more attention to their methods and to document them explicitly. The watershed for the birth of modern documentary editing has often been placed at the Middle English editions of Frederic Madden (1801–1873), especially his third one, the 1839 edition of the *Syr Gawayne* anthology where he found a middle ground between the more extreme editorial practices of his two earlier editions.<sup>41</sup>

One problem in defining the principles and practices of documentary or non-critical editing is the fact that the idea of a ‘faithful’ or ‘diplomatic’ transcription seems to vary widely among editors, especially in the 19<sup>th</sup> and early 20<sup>th</sup> centuries, as does its evaluation by later scholars. For example Well’s (1907) edition of *The Owl and the Nightingale* claims that “[t]he MSS. are printed as much in facsimile as possible’ with any reading that makes sense left as it stands”, but states that abbreviations are italicized, and punctuation, capitalization and word division are editorial (Blake 1998: 69). Especially problematic in this respect are the numerous and influential editions of the EETS, which vary in their editorial approach and have also been variously evaluated by later scholars.<sup>42</sup> For example, while Machan (1994) sees the approach of the EETS as diplomatic and lexical, Edwards (2000: 74) sees it as more of a counterpoint to Madden’s more documentary approach. The evaluation of the tradition of Middle English editing seems to depend heavily on the viewpoint of the evaluator, as editors of a critical persuasion tend to see it as exceedingly conservative in orientation, while those of a more diplomatic persuasion consider it to be overly critical and eclectic in its approach.

Up to this point the kind of editing discussed here has been called ‘non-critical’, ‘diplomatic’ or ‘conservative’ editing. The term *documentary editing* adopted here became current in the United States in the late 1970s—when the methodology itself had been in use for over two decades—and it was used to distinguish this editorial technique “from the more traditional approach of textual editors,

<sup>41</sup> The two earlier editions were of *Havelok the Dane*, published in 1828, and of *William and the Werewolf*, published in 1832. Of these, the former made more concessions to modern typographical conventions while the latter was a strict type facsimile, printed in blackletter type. In the *Syr Gawayne* edition Madden “adopts a quasi-facsimile procedure, abandoning black letter and signalling final flourishes, barred -lls and representing contractions typographically”, and documents all of his occasional emendations (Edwards 1987: 44-5, 2000: 72).

<sup>42</sup> The society’s current editorial principles (EETS Guidelines) point out that no “prescriptive set of editorial principles is laid down, but it is usually expected that the evidence of all relevant medieval copies of the text(s) in question will have been considered”, which gives individual editors considerable leeway, allowing equally conservative parallel-text editions and critical editions that construct a single text out of a number of witnesses.

who consciously applied critical judgement and scholarly experience to produce new, editorially emended texts for their audiences” (Kline and Holbrook Perdue 2008: 2), which was found to be inappropriate for the editing of the original papers of statesmen and other significant personages. The documentary approach was adopted—with slight methodological and presentational differences—by editors of both “papers of public importance” and the private, unpublished papers of literary authors, which were “viewed as expressions of a person’s private feelings and internal development” (12). While editors of the papers of great individuals—whether statesmen or literary authors—habitually emended and standardized the linguistic features of the papers to present a uniform style, editors who in the 1970s began editing the records of groups and organizations quickly found that they could not define a uniform style to standardize to, and thus had to “adopt far more literal methods of presenting editorial texts”, contributing to “a silent revolution of textual methodology among historian-editors” (16).

In 1978 G. Thomas Tanselle published an essay, titled “Editing of Historical Documents”, which “irrevocably changed the nature of American documentary editing” by taking post-World War II documentary editors to task on two scores. First of all, it argued that “the statements of textual method in these volumes were often maddeningly vague and occasionally self-contradictory” and second, that “the application of heavily emended transcription instead of more conservative methods of literal transcription was a disservice both to the documentary sources and to their readers” (Kline and Holbrook Perdue 2008: 19). Although there was still disagreement on the level to which editorial emendation should be documented, there emerged in the 1980s and 1990s a realization that editors of different documentary persuasions “shared the same goal in dealing with documentary sources: how best to present cautiously emended texts that preserved as much as possible of the original’s evidence” (Kline and Holbrook Perdue 2008: 25). As Williams and Abbott (1999) have noted, this has meant that historians have over the last two decades followed the trend mentioned at the beginning of this chapter and “become more aware of their need for soundly edited texts, often in documentary editions”, instead of the traditional kinds of editions of historically interesting documents, which were “often edited with more concern for their informational than for their evidentiary value” (Williams and Abbott 1999: 71-2).

### 3.2.1 Theory and methodology of documentary editing

Documentary editing has frequently been seen as largely untheoretical (McGann 2004: 380), and both its theoretical underpinnings and practices have certainly received less attention than those of critical editing. Kline and Holbrook Perdue (2008) define the task of a documentary editor as the translation of “handwritten, typescript, or printed source texts into a form that their readers can trust as an accurate representation of the specific original materials they represent” (143). This means reproducing the wording, spelling, punctuation, etc. of the textual content of the original document as accurately as possible, since unnecessary emendation will make the editorial text of any document “useless as evidence either of what its author wrote or of what its addressee read” (130).<sup>43</sup> Thus, unlike critical editing

<sup>43</sup> In this context, the meaning of ‘necessary’ and ‘unnecessary’ has been defined in terms of the practical limitations of modern print technology, as is explicit in the “classic definition” of a diplomatic

which focuses on recovering the *work* behind the various surviving *documents*, documentary editing is—as befits its name—interested in representing the *document* or the “historical artifact” (Williams and Abbott 1999: 71).<sup>44</sup>

Although a documentary or diplomatic edition represents the specific features of an individual document (or a series of documents, as will be shown below), it should not be seen merely as a synonym for a diplomatic transcription (a view implied e.g. by Gabler 2007). Although a documentary edition is always based on a diplomatic transcription of the original document, these two concepts identify two very different objects representing separate stages of the editorial process: while the transcription is “a derivative document that holds a relationship with the transcribed document”, the diplomatic edition constitutes “a formal (public) presentation of such a derivative document” (Pierazzo 2011: 464). This distinction between mere replication and analytic ‘re-representation’ also lies at the heart of the claim of documentary editions for the status of *scholarly editions*. As Lavagnino (2006) has emphasized, it is the task of documentary editors “to make specific claims about what the text is and communicate them clearly to others” (338).

According to Kline and Holbrook Perdue (2008: 36-7), documentary editors have three basic goals:

- 1) to create a verified, trustworthy text that can be read by modern audiences;
- 2) to make documents more available to a wider audience than the small group of people who might be able to view originals in their home archives; and
- 3) to provide contextual aids that make the documents more readily understandable by offering readers the “historical, literary or technical context in which to make the best use of them”.

As these aims illustrate, documentary editions should be “properly regarded not as the end of scholarly research but as its beginning” (Kline and Holbrook Perdue 2008: 289). Unlike the critical editor, who is supposed to have exhausted the significance of the evidence in preparing his own interpretation of the text, the documentary editor is not merely presenting the fruits of his own research for the edification of the scholarly community, but rather establishing “a foundation of evidence on which others will build” (289). This concept of the edition as a research tool or a *data archive* to be used as the starting point for research is central to this thesis and will be discussed in more detail in ?? and chapter 5.

---

edition presented by Pierazzo (2011), being “a transcription that reproduces as many characteristics of the transcribed document (the diploma) as allowed by the characters used in modern print”, including “features like line breaks, page breaks, abbreviations, and differentiated letter shapes” (463-4).

<sup>44</sup> Here it is important to note the choice of words by Kline and Holbrook Perdue: while documentary editors take as their starting point a particular *document*, they have traditionally been interested in the *text* “as it was available at a particular time in a particular document” (Williams and Abbott 1999: 71), with only limited attention to the paratextual aspects of the document. I argue that just as in the case of critical methodology, this has been at least partly due to the practical constraints of print, and that the development of digital technology over the last two decades has removed (or at least mitigated) many of the reasons for restricting our attention exclusively to the textual aspects of historical documents (see chapter 4).

### Basic assumptions of documentary editing

In contrast to textual critical editors and like Jerome McGann, documentary editors of medieval texts “proceed on the assumption that notions of authorship are of secondary importance in medieval literary culture” (Robins 2004: 150). While the validity of this assumption might be debatable in the case of literary texts, it is in fact the only reasonable assumption in the case of anonymous utilitarian texts like recipes, as was established in subsection 3.1.2 above. Instead of linking the textual object to its author, documentary editors have traditionally linked it “either to its historical audiences or to the material practices of the period of its first production” (Eggert 2004: 165), moving the focus away from authorial intention and towards the transmission process of the text (see e.g. Roland 2004 and Klinck 2004):

Critics who reject the intentional fallacy and have a healthy respect for history tend to think of texts as autonomous. One logical consequence of this attitude is to prefer texts with historical integrity, artifacts from the past meaning whatever the words on the page seem capable of meaning regardless of any intending consciousness behind them. (Shillingsburg 1986: 13)

Another central assumption in the documentary editing of medieval manuscripts is that each manuscript document is “a unique production” (Fellows 1998: 15), and a medieval reader would have been unlikely to have access to more than a single version, simply having to “make the best of whatever text he or she could get a hold of” (Blake 1998: 68). Thus documentary editors consider it “more pertinent to an understanding of medieval culture to concentrate on what was actually read than to pursue the elusive chimaera of original authorial intention” (15):<sup>45</sup>

We are interested in documentary evidence precisely because it encodes, however cryptically at times, the evidence of the agents who were involved in making and transmitting the document. Scholars do not edit self-identical texts. They reconstruct a complex documentary record of textual makings and remakings, in which their own scholarly work directly participates. (Buzzetti and McGann 2006: 71)

As Shillingsburg (1999) points out, the increased popularity of more documentary approaches to editing is related to the recognition in the 1990s of “the textual implications of the bibliographic entities in which readers encounter linguistic texts” (63), and to the social textual criticism of McGann. When “texts as historical documents” are seen as “the material results of complex forces—including the author but not exclusively focused on authorship”, they naturally become “the proper focus for students of texts” (63) (Flanders 2006: 142).

<sup>45</sup> While it is of course impossible—and for most purposes also beside the point—for an edition to try and recreate a medieval reading experience (see e.g. Shillingsburg 1986: 12) because of the differences in the surrounding context, Roland (2004) argues that a “documentary edition, be it a parallel-text or single-text edition, provides a text that approximates, lexically though not visually, a historical document—that is, a text that was actually read” (42).

### Relationship to critical editing

These assumptions and requirements naturally result in a very different kind of editorial ideal than that of critical editing. This has often provoked accusations that documentary editors “in effect cease to edit” (Jacobs 1998: 5). This is not necessarily untrue, but neither is it necessarily a bad thing, as Kline and Holbrook Perdue (2008) point out, quoting a lecture by Steven Meats to editors of literary correspondence:

As a general rule, the less editing (that is, the less editorial emending and altering) done to the text of letters, the better the job of editing. A letter is, after all, a primary historical document; one might even call it a ‘fact’. In any case, silent emendation in the editing of letters should be severely restricted. (Kline and Holbrook Perdue 2008: 130)

Although Meats is here referring specifically to letters, Kline and Holbrook Perdue (2008) emphasize that this model of ‘minimalist editing’ is a generally accepted tenet of documentary editing, pointing out that “‘cautious’ and ‘conservative’ [...] are now the accepted watchwords among documentary editors for considering all textual matters” (209). This editorial ethos is in sharp contrast to the attitudes represented by traditional critical editing and has at times attracted rather savage criticism from critical editors. However, editors and scholars with experience of documentary and other ‘non-critical’ editing are very much unanimous that non-critical editions “should not pretend to be noneditorial”, as “the editor is always present in the organization of the material and the transcription of source documents” (Vanhoutte 2006: 164), as well as in the very decision determining which historically authoritative document gets reproduced in the first place (Shillingsburg 1986: 84-5):

Documentary editing, although noncritical in terms of classical textual scholarship, is hardly an *uncritical* endeavor. It demands as much intelligence, insight, and hard work as its critical counterpart, combined with a passionate determination to preserve for modern readers the nuances of evidence. (Kline and Holbrook Perdue 2008: 3)

While Kline and Holbrook Perdue (2008: 125) argue that documentary editing makes no distinction between printed and manuscript materials, there has historically been a rather clear division of labour between the critical and documentary approaches based on the type of text to be edited, with exts that were “never intended for publication—such as journals, notebooks, and letters” (Cohen and Jackson 1991: 106) having been edited following the documentary approach, and published works having been edited critically (Cohen and Jackson 1991: 106; Williams and Abbott 1999: 71).<sup>46</sup> While the criterion of publication, associated with printed texts, is problematic in the case of medieval manuscripts, they were nevertheless vehicles of communication intended for public dissemination, unlike most modern manuscripts which are mostly unique and private in nature Pierazzo (2009: 179).

<sup>46</sup> In practice, the texts edited in a documentary fashion can in most cases be considered *unique*, which effectively precludes those critical approaches that are based on the comparison of multiple surviving witnesses (Rosenberg 2006: 98).

Also in the case of medieval texts, unique *documents* that are considered historically significant in themselves have usually been edited following documentary principles, while established literary *works* surviving in a large number of manuscripts and considered artistically or culturally significant have been edited critically. Texts falling between these extremes—practical or utilitarian texts, popular literature or other kinds of ‘everyday writing’—have received much less editorial attention and their editorial treatment has not been as firmly established, vacillating between conservative documentary editions of individual manuscripts to critical editions of larger manuscript traditions. In the more specific context of medieval recipe collections—discussed in more detail in chapter 8—the editorial approaches adopted have ranged from the traditionally critical—even eclectic—edition of a number of Middle English collections by Hieatt and Butler (1985) to Terence Scully’s quite conservative parallel and single text editions of the French *Viandier de Taillevent* (Scully 1988) and *Vivendier* (Scully 1997) collections, and the consecutive presentation of all four manuscript texts of the *Libellus de arte coquinaria* by Hieatt and Grewe (Hieatt and Grewe 2001).<sup>47</sup>

### Relationship to facsimile editions

The only objective thing is the unique original manuscript itself; it may not be replaced by an equivalent, and in the strict sense may not be reproduced even by color photography. The material manuscript itself, not the text of the manuscript, is the record. The manuscript requires interpretation, however, and the result of the interpretation is the text.  
(Zeller 1995a: 43)

Critical editors have often seen documentary editions merely as more easily printable and readable versions of facsimile editions, serving largely the same function (see e.g. Kiernan 2006), and some have even claimed the digital facsimile to have superseded the diplomatic edition (Allen 1987: 12). Although the creation of digital facsimiles has become much easier and cheaper than in the 1980s, this supersession does not seem to have taken place. While extremely useful for palaeography and usable for many types of literary or historical scholarship, photographic facsimiles are not very useful for historical linguists who have increasingly moved towards corpus-based methodologies requiring searchable, digitally encoded texts. Instead of superseding it, a facsimile presented alongside an editorial transcription of the text can supplement it in two ways: First of all, it can serve as a record against which the interpretation represented by the transcription can be verified (Zeller 1995a: 43), and second, it can provide access to those documentary features that cannot be represented by the transcription but nevertheless an important factor in the construction of documentary meaning, i.e. its *material paratext* (McGann 1991: 12, 149).

Apart from the most obvious reason—that a manuscript text presented in facsimile is not any easier for the untrained reader to read (Haugen 2004: 76)—the photographic facsimile has not replaced the documentary edition as a means of editing manuscript documents because it does not “articulate it in the way the

<sup>47</sup> The relative chronology of these editions also illustrates the historical movement towards more documentary editions since the 1990s.

typographical presentation does” (Zeller 1995a: 47-8). This makes the facsimile an alluring but unfortunate compromise; neither an adequate replacement for the original, nor an analytical operation performed on the original to bring out what is relevant, like a documentary edition based on transcription (Cerquiglini 1999: 21-2). Although sometimes placed in the same category, a documentary edition is fundamentally different from a facsimile edition, being the result of a “non-objective, interpretative operation” and presenting a “scholarly and computational analysis of the chosen textual phenomena” (Pierazzo 2011: 472) instead of a mere mechanically produced visual surrogate.<sup>48</sup>

It is precisely this *interpretive* nature of transcription that provides even the most verbatim documentary edition with its status as a scholarly product. For example Greetham (1987: 64), Zeller (1995a: 19) and Pierazzo (2011: 466) have observed that the editor inevitably interprets already when deciphering the manuscripts, and the way in which the editor organizes an edition strongly influences the overall picture that we get of the texts, making it a “truism that there is no such moral security as a perfect loyalty to a document” (64).<sup>49</sup>

### Practical methods of documentary editing

Even in documentary editions, “the patterns of characters, words, phrases, and paragraphs offered to the reader are seldom the only ones that the edition’s source could have produced”, but rather the result of the interaction of the chosen editorial methods with the source (Kline and Holbrook Perdue 2008: 141). Whereas the methodological apparatus of critical editing is focused on the synthesis of a single text from multiple witnesses, documentary editing is based on the premise that a documentary edition should contain “the words, phrases and punctuation of a single source that should be readily and conveniently available to the reading audience” (Kline and Holbrook Perdue 2008: 87). While this eliminates the need for selecting the ‘correct’ option from available local variants, it does not do away with the element of choice but merely moves it up to the level of the entire *document* (Kline and Holbrook Perdue 2008: 87).

After a documentary editor has made the crucial decision of choosing her source document, she “must cautiously move from the ‘original,’ the document that holds evidentiary value, toward some accessible version that will serve the needs of a wide audience” (Kline and Holbrook Perdue 2008: 109), i.e. to transcribe it into a new, more accessible medium—usually a printed book or digital encoding. Although documentary editing is usually associated with verbatim transcription, the range of presentational approaches taken by documentary editors extends from photographic facsimiles to what amounts to an eclectic text silently incorporating all scribal corrections made to the document. Kline and Holbrook Perdue (2008) have outlined five general methodological frameworks within which documentary editing can take place, resulting in different kinds of editions:

<sup>48</sup> Fortunately, modern digital media not only make the use of photographic facsimiles easier, but they also allow the analytic representation of many paratextual features as an integral part of a diplomatic transcription (see section 5.4 and subsection 10.1.2).

<sup>49</sup> Even Lass (2004b), adamant in his requirement for absolute fidelity to the source text, recognizes that not even “the most careful and expert diplomatic transcription will yield a text ‘wie es eigentlich war’” (22).

- 1) photographic and typographic facsimiles,<sup>50</sup>
- 2) editorial texts requiring symbols or textual annotation,
- 3) diplomatic transcriptions,
- 4) ‘inclusive texts’ and ‘expanded transcriptions’, and
- 5) clear text.

Of these options, listed in order of decreasing fidelity to the original document, facsimile editions have already been discussed above and found to be fundamentally different from transcriptions and ill-suited for the purposes of historical linguistics. Clear text editions are essentially similar to silently emended best-text editions and share the same problems, being “misleadingly smooth” (Zeller 1995a: 40) and suppressing “many elements of inscription, making recovery of the details of the original extraordinarily difficult” (Kline and Holbrook Perdue 2008: 146).<sup>51</sup> Thus, it is the middle three—along with the typographic facsimile—that are the most relevant with regard to the present edition.

According to the traditional definition, a typographic facsimile “attempts to duplicate exactly the appearance of the original source text as far as possible within the limits of modern typesetting technology” (Kline and Holbrook Perdue 2008: 147-8).<sup>52</sup> In addition to representing the text of the document, the type facsimile also tries to provide information “about the physical state of the manuscript and/or the various levels of revision it may reflect” (Edwards and Moffat 1998: 235). In print editions, typographic facsimiles have been used mostly for reproducing *printed* documentary sources (Kline and Holbrook Perdue 2008: 149), perhaps because the similarity of the source and target media makes faithful representation relatively easy. However, modern digital encoding and annotation technologies make it relatively easy to produce detailed type facsimiles of even manuscript texts and combine their visual expressiveness with more analytical forms of transcription.<sup>53</sup> The second item on the list, editorial text requiring symbols or textual annotation, is not really a separate ‘level’ of editing, but simply refers to editions of types 3–5 that use a variety of textual notes and editorial symbols to represent para- and metatextual elements in the original—such as layout (line and page breaks, etc.), passages deleted or added, damage or gaps in text and special symbols (Kline and Holbrook Perdue 2008: 152-61).

The third method—the diplomatic transcription—has been used for a wide variety of document types from modern letters to ancient papyri. Kline and Holbrook Perdue (2008) define it as referring to an edition, where the details of inscription

<sup>50</sup> Considering the fundamental differences between photographic and transcriptional representation discussed above, the concatenation of photographic and typographic facsimiles is in fact quite misleading as they belong to different sides of the analytical divide.

<sup>51</sup> Clear text editions are used mainly in cases where the original document is either so heavily revised by the author or scribe that it essentially contains several versions of the text in the same document, in which case it is essentially a critical edition, or so clear and unproblematic that no emendation or special annotation is required (Kline and Holbrook Perdue 2008: 173-4).

<sup>52</sup> In practice this means that it “reproduces the author’s spelling and punctuation without any correction”, does not expand “contractions or abbreviations”, prints the “author’s additions above the line [...] interlinearly”, sets marginal notes in the margins, represents passages “crossed out by a line [...] in canceled type”, follows the author’s layout exactly, and sets “[h]eadings, titles, datelines, greetings, and salutations [...] line for line so that the line breaks in the printed version mirror those in the original” (Kline and Holbrook Perdue 2008: 147-8).

<sup>53</sup> While the present edition is not intended to be a conventional type facsimile, it does incorporate several features traditionally associated with them (see chapters 10 and 11).



are indicated using critical symbols or abbreviations instead of being represented visually, the placement of “routine elements”<sup>54</sup> is often standardized, “[m]issing punctuation” added, “ambiguous or archaic abbreviations and contractions” (161) expanded, and “words unintentionally omitted by the author or destroyed by the mutilation of the original” (161) supplied. It should, however, be noted that the term *diplomatic edition* has been used in a number of different editorial traditions with slightly different definitions.<sup>55</sup>

One definition that quite accurately describes the practice of diplomatic transcription as it is understood and practised in the present edition was presented by Rhys and Evans already in 1887, differentiating the diplomatic transcript from the typographic facsimile based on the level of transcriptional detail:

A diplomatic reproduction differs from a facsimile chiefly in one particular - it does not profess to give the special form of the manuscript characters, but it should give character for character, letter for letter, word for word, *spacing for spacing*, error for error, deletion for deletion, correction for correction, rubric for rubric; in short, there must be no tampering of any kind, not even with the punctuation.  
(Rhys and Evans 1887: xiv)

In other words, while type facsimiles operate on the *graphetic* level, distinguishing between different graphetic forms of the same *grapheme* (long and short ⟨s⟩, single and double compartment ⟨a⟩, etc.), diplomatic transcription is *graphemic* in nature, abstracting the specific letter-shapes into a typology of distinct *graphemes*, usually defined as the smallest semantically distinguishing unit in written language (analogous to a *phoneme* in spoken language).<sup>56</sup>

The *inclusive text* or *expanded transcription* is described by Kline and Holbrook Perdue (2008: 163-73) as a middle ground between a detailed diplomatic edition and a clear text edition.<sup>57</sup> Both of these approaches provide a lightly standardized and emended transcription of the original document, where some details of inscription (such as additions, deletions, corrections, etc.) are reported in the text itself using either textual symbols or numbered footnotes, and some are relegated to textual notes at the back of the book.<sup>58</sup>

<sup>54</sup> These include things like “datelines, greetings, salutations, titles and the indentation of paragraphs” (161).

<sup>55</sup> In the French editorial tradition, Foulet and Speer (1979) define a diplomatic edition in its strict sense as “a printed representation of all the particulars of a manuscript text, without benefit of such adjustments as the separation of words run together and the addition of diacritical marks and modern punctuation”, where even “abbreviations, reproduced in type as accurately as possible, remain unexpanded” (43).

<sup>56</sup> The transcriptional practices adopted in the present edition and the issue of graphemic and graphetic transcription are discussed in more detail in section 10.2.

<sup>57</sup> The difference between the two, according to Kline and Holbrook Perdue (2008: 163), is not in their degree of conservatism or liberalism in emendation—which can vary in both—but rather in the degree to which editorial intervention is reported to the reader, inclusive texts (associated with the Center for Editions of American Authors (CEAA) and the Committee on Scholarly Editions of the Modern Language Association and editing of the documents of literary authors) being more verbose in this regard than expanded transcriptions (associated with editing of the documents of historical figures).

<sup>58</sup> In the French editorial tradition, this kind of edition is known as an “interpretive diplomatic edition” (Foulet and Speer 1979: 44), which is seen to be a step towards a critical edition from a pure diplomatic edition, with the words separated, abbreviations expanded in italics and occasionally

In terms of this fivefold categorization, the present edition incorporates features from all of the four edition types discussed above. While its basic transcriptional strategy is that of a diplomatic edition, it also includes data about many of the visual features traditionally reproduced only in type facsimiles and uses inline annotation within the text to record layout, scribal emendation and damage. However, due to its digital format, it is not fixed to any of the abovementioned presentational formats, but allows the edited text to be viewed equally as a restricted type facsimile, a variously annotated diplomatic transcription and an ‘expanded transcription’, depending on the needs of the user.<sup>59</sup>

### 3.2.2 Criticism of documentary editing

For every scholar at work on producing a ‘documentary’ textual base that will allow others to produce their own critical texts, there is another who complains that this is an abandonment of scholarly responsibility, a refusal to point out to readers what is ‘better’ or ‘worse’, ‘significant’ or ‘irrelevant’. (Kline and Holbrook Perdue 2008: 29)

Much of the criticism presented against documentary editing could be called ideological in nature, being based not on the practical shortcomings of documentary editions or their unsuitability to their intended purpose, but rather on their failure to fulfil what is perceived as the ‘proper editorial function’.<sup>60</sup> Documentary editions of a single manuscript witness have traditionally been seen as a last resort, used in situations where critical editing is not possible: “faced with a particularly inscrutable or badly damaged text, editors may defensibly conclude that a purely transcriptional, that is, diplomatic, edition is their only option” (Moffat 1998: 35-6).<sup>61</sup> Many editors and textual critics seem to be fixated on the necessity of editorial intervention, as if it was necessary for justifying the efforts of the editor.<sup>62</sup> The reluctance of editors to intervene in the text of their source documents

even modern punctuation added. Haugen (2004), however, objects to the use of this term, preferring to call it simply “diplomatic transcription”, because the former term would imply “that there are diplomatic editions without interpretation” (79).

<sup>59</sup> In addition to the edition as a *data archive* from which different presentations can be generated, appendices B, C and D include three different presentations of this data: a detailed diplomatic rendition of each page, an annotated clear text reading edition with all parallel versions of each recipe presented side by side, and a browsable digital presentation combining the two.

<sup>60</sup> The underlying idea in much of the criticism directed at non-critical editing seems to revolve around the idea that it involves no intellectual or scholarly value. As an extreme version of this prejudice, Ore (2004) mentions a conference paper given in 1998 (Ore 1999), where “[o]ne of the comments in the discussion after the presentation was that I was suggesting the kind of textual criticism that could be performed by trained monkeys, that unless an editorial project included text selection and emendation, without the critical edition, there was no work of scholarly value” (Ore 2004: 39).

<sup>61</sup> For example Keiser (1998c) argues that with the availability of various tools—like library catalogues and indices like the *Index of Middle English Verse* and the *Index of Middle English Prose*—for locating different manuscript versions of the text, “the justification for producing diplomatic or semi-diplomatic editions based on only one manuscript is far less than it was in the recent past” (111), as if the inability to locate more versions was the only reason for producing a single-witness edition.

<sup>62</sup> In discussing the editing of texts surviving in a single copy, Moffat (1998) writes that the “editor of a work found in a unique copy must employ other strategies [than those of textual criticism] and justify policies of intervention or nonintervention on other theoretical bases” (35), as if intervention was a *sine qua non*, for which the editor must find some pretext.

has frequently been criticized as “pusillanimity” (Edwards 1998a: 100), non-critical editions being seen rather condescendingly as “over-conservative”, representing merely “the shape represented by one of the manuscripts, timorously collated against a few of its fellows and occasionally corrected” (Hussey 1992: 106, quoting George Russell).<sup>63</sup>

### Representation of a single manuscript

The dismissal of these ideological criticisms as the result of text-critical prejudices and assumptions about the editorial role does not mean that the theoretical basis and practices of documentary editing are above criticism. One frequent target of criticism, raised for example by Machan (1994: 184), has been the claim—sometimes presented as a justification for documentary editions—that a diplomatic edition of a single manuscript text represents what at least one medieval reader read. However, as Machan (1994) points out, medieval readers “read by supplementing the text in front of them from a variety of cultural and literary networks” (184), which are not available to a modern reader, making her experience fundamentally different from that of the medieval reader. This has led Machan to argue that “to produce a properly historical document, one must attempt to recover what lies behind the individual manuscript” (Echard and Partridge 2004a: xv), supplementing the documentary text with information that tries to give the modern user of the edition some idea of the cultural and situational context of the medieval reader. While I completely agree with Machan’s criticism, this observation applies equally to the original document itself, and is thus not a criticism of documentary editing as such, but simply of the New Critical view of the textual object as a container of objective meaning. In terms of the model presented in chapter 2, Machan’s observation means that since the situational and cultural contexts of the medieval and modern readers are radically different, the *works*—or the semantic interpretations of the text—constructed by them from the same document are inevitably different.<sup>64</sup>

Since a *linguistic* approach to editing is not primarily interested in the *work* as constructed by a reader, but rather on the *document* as a material witness to a linguistic act performed by the *writer*, the function of a ‘linguistic’ edition is also radically different. It is not intended to reproduce or even simulate the medieval reading experience, but rather to allow the study of the linguistic act of the writer (i.e. scribe) through the analysis of the textual and paratextual evidence provided by the original document. The edition, thus defined, is a *selective analytical model*<sup>65</sup> of those textual and paratextual features of the original document that are considered to be potentially relevant for the interpretation of that linguistic act from a given theoretical viewpoint. And for this purpose, as will be argued in the

<sup>63</sup> For example Cooper (1998) makes a rather disparaging remark about N.F. Blake’s faithful edition of the Hengwrt MS of *The Canterbury Tales*, saying that “he does not so much edit as transcribe the text” (88).

<sup>64</sup> Also the *text* itself that the two readers construct on the basis of the document can already be different due to the differing linguistic context, the most obvious example of a discrepancy being the potential inability of the modern reader to correctly interpret its medieval abbreviations.

<sup>65</sup> This formulation also acknowledges the fact that any act of reproduction is always selective and that the very act of reproducing the text in a different medium—“be it script, typeface, or hypertext”—necessarily alters the readers’ perception of the text (Roland 2004: 52).

next chapter, a detailed documentary edition is the most effective tool.

Hanna (1987) and Allen (1987) have also criticized single-text editions for being arbitrary and eclectic in uncritically accepting all the decisions of a single scribe, arguing that while the scribal mind might be 500 years closer to the original than the editorial one, this “does not diminish the fact of that mind’s uniqueness, nor does it vouch for its taste and skill at every potential lemma” (Hanna 1987: 89–90). According to them, this means that trusting a single medieval scribe is no different from trusting a modern editor, neither being the original author. This argument, based entirely on the text-critical fetishization of authorial intention, is absolutely irrelevant from a linguistic point of view. While a historical linguist is generally just as interested in the linguistic performance and judgement of the medieval scribe as of the original author, she has very little interest in the linguistic performance and judgement of the modern editor.

A more pertinent criticism of the documentary practice of using a single manuscript to represent the entire textual tradition is the claim that it distorts the variant nature of medieval textuality by suppressing the other versions and the variation implied by them. This means that the single documentary version chosen by the editor not only comes to determine “all future perceptions of the work in question” (Sturges 1991, quoted in Edwards 1998a: 101), equating it with the specific features of the particular documentary version,<sup>66</sup> but—more importantly for historical linguistics where the concept of both synchronic and diachronic *variation* is key—it would also obscure the degree and kind of linguistic and textual variation between the surviving witnesses and provide an excessively stable view of both medieval textuality and the historical linguistic situation. However, as will be pointed out in subsection 3.2.3, even these problems can be effectively solved by a documentary edition including and juxtaposing all relevant manuscript versions of the work.

### Inclusion and presentation of data

Although the theoretical basis of documentary editing and its basic premise of producing an accurate representation of a single document are highly compatible with the needs of historical linguistics, the practical realization of this aim and the methods employed have often been somewhat problematic. While documentary editors have historically been criticized for their lack of editorial intervention and providing too much raw data,<sup>67</sup> I would—from a historical linguist’s point of view—argue the opposite: documentary editions have in fact neglected to live up to the full measure of their principles, making too many compromises in the face of critical pressure. Even documentary editors who claim to provide a diplomatic

<sup>66</sup> Despite pointing out the problem, Sturges also observes that since “that version is an actual medieval one, the charge of falsification or unconscious modernization is greatly reduced”, making documentary editing of a single manuscript version “preferable to critical editing for medieval works” (Sturges 1991, quoted in Edwards 1998a: 101).

<sup>67</sup> For example, editions of the papers of literary authors published in the 1960s in accordance with the guidelines of the CEAA were criticized for mixing “the important with the inconsequential” (Kline and Holbrook Perdue 2008: 14) since they print the journals and papers of authors in their entirety. This problem was seen to be exacerbated by the addition of too much descriptive annotation into the text itself, leaving the reader “stumbling over scholarly roadblocks and barricades” (14) and making it even more difficult to focus on what is essential.

transcription of the source document usually end the description of their editorial principles with a caveat listing the various kinds of standardizations and normalizations that have been performed on the text as standard practice.<sup>68</sup>

These editorial interventions in the text, like the expansion of abbreviations, supply of editorial punctuation, correction of ‘obvious textual faults’ or standardization of the layout always carry the risk of obscuring or falsifying linguistically significant features of the original document.<sup>69</sup> A much better alternative—and much truer to the stated principles of a diplomatic documentary edition—would be to reverse the situation, providing a true diplomatic transcription of “what is physically present in the source” and use the editorial annotations to indicate the editorial interpretation of “what ought to have been there but for one reason or another isn’t” (Driscoll 2006: 260). This simple shift in polarity would provide the historical linguist with a valid record of “the historical document, warts and all” (Shillingsburg 1986: 21), fulfilling the requirements for a *linguistic edition* discussed in section 4.3 while still providing the historically or literarily minded reader with the necessary aids to make sense of the text.

### 3.2.3 Documentary editing of multiple versions

One of the main limitations of the documentary approach to editing is its focus on unique documents. This is naturally not a problem in the case of documents such as letters, diplomas and private journals that have only ever existed as a single copy, or in the case of texts with a linear evolution culminating in a ‘finished’ (which usually means ‘published’) version. But as Zeller pointed out already in 1975, texts—especially manuscript ones—frequently exist in several versions with often complex and indeterminate relationships. As was observed above, the usual solution for such texts has been to attempt the reconstruction of their archetypal forms through text-critical methods, which was already shown to be a method unsuitable for both medieval manuscripts and for editions intended for linguistic research tools. The other option discussed so far, i.e. a documentary edition of a single ‘representative’ document version of the work is equally problematic for reasons listed above.

The truism that ‘every medieval manuscript is unique’ is especially applicable to Middle English scientific and practical texts which were often quite freely adapted by their scribes, as was pointed out in subsection 3.1.2. The examination of the various versions of a work in surviving manuscripts can therefore provide information about the “contemporary reception and understanding” of the work, and thus on its nature “and the uses that its compiler intended it to have” (Keiser 1998c: 113). This makes the representation of textual variation between the surviving manuscript versions a priority for the editor, since it can offer information not just about the textual transmission of the work itself, but also about the perceptions, attitudes and valuations of the society and culture in which it was

<sup>68</sup> Assuming that they explicate their editorial principles and practices at all, which—according to Kline and Holbrook Perdue (2008) is not all that common for documentary editors.

<sup>69</sup> Even explicit documentation of these changes—either with special symbols or textual notes—is a small comfort for the linguist wishing to use the text as a part of a corpus, because they mean the he or she either has to in essence re-edit the text, reversing the interventions of the editor based on the documentation or suffer their consequences.

transmitted.<sup>70</sup>

For these reasons many editors and critics have over the last twenty or so years admitted the fact that a single established text is perhaps not the most useful product for an editorial project:

Some editors seem to think that a text representing an ultimate intention is the goal of editing. I contend such a goal ignores what we know about the fluctuation of intention, and that literary works of art, unlike some other forms of art, cannot safely be treated as single end products. Critics, in short, should demand from editors something more than a single, simple authenticated text.

(Shillingsburg 1986: 41)

Fortunately, the production of multi-text editions of medieval texts has since become an accepted practice, being occasionally even presented as the preferable method. For example McGillivray (1994) is of the opinion that the “actual state of a medieval work can be presented with the most openness and honesty by publishing all of the manuscript versions of the work in a way that allows the reader to compare the various readings as they exist in the different texts” (184), while Johansson (2004) argues that “all primary sources, i.e. all witnesses of e.g. an Old Norse text, are relevant to researchers in different disciplines, and therefore should be made available when we work with electronic tools” (93-4).<sup>71</sup>

The inclusion of all manuscript versions of a work (either in facsimile or diplomatic transcription) has also been seen to have specific practical benefits for the editor. For example McGillivray (1994: 187-8) has argued that it frees the editor from two detrimental imperatives: “the necessity of always being right and the necessity of always providing a critical text” (187).<sup>72</sup> It has also been seen as “highly desirable” for the reason that each version helps to cast light on the other versions, and the opportunity to study related texts is thus “greatly enhanced by having the related texts available within the same book” (Keiser 1998c: 114-5). From a linguistic point of view, reproducing several versions of a text also has the additional virtue of representing two forms of English prose from different time periods but with similar content (Keiser 1998c: 114). Although the promotion of documentary multi-text editions has been more common among editors of non-literary texts, their relative merits versus eclectic clear text editions have been

<sup>70</sup> Furthermore, as Doyle (2000) has pointed out, this kind of information is not encoded solely in *textual* variation between the different manuscript versions, but also in the *material paratext* of the documents: “The more manuscripts of the same text one examines in comparative study the more one is challenged to understand why they differ or accord in certain physical respects. And to ask, can such physical details tell us more about the makers and circumstances of making?” (Doyle 2000: 7)

<sup>71</sup> Also more traditional editors have expressed a more qualified acceptance of a multi-text edition as an alternative to critical editions in special cases (see e.g. Fellows 1998: 22-3).

<sup>72</sup> The first of these imperatives has the effect of forcing the editor toward conservatism: when the edited text is all there is, the editor cannot afford to take risks or conjecture, but if the edited text is backed up by the full texts of all witnesses, the conjectural nature of the edited text is more explicit and the editor could afford to take greater risks. The second imperative is a specific variation of the first: when the edited text is all there is, the editor has to choose one of the readings even if none of them are likely to be authorial, so as not to leave a gap, but if the reader has recourse to all the manuscript versions, the lack of a likely authorial reading can be indicated by leaving the gap visible.

debated also among literary editors.<sup>73</sup>

### Selective representation of multiple texts

Editors and textual scholars have suggested a variety of practical solutions for the diplomatic representation of manuscript texts surviving in multiple copies. Perhaps the simplest solution, recommended by Moffat (1998: 42) for two texts that represent either “obviously related works” or clearly different versions of the same work, is to edit them as two separate editions. In cases where the text survives in a large number of manuscript versions and it is not feasible to represent all of them, Robins (2004: 150) has suggested the presentation of a single-text edition of a version that is considered ‘central’ to the textual tradition in juxtaposition with a similar edition of a version that is considered peripheral. This kind of selective editing, whose purpose is to ‘destabilize’ the text and remind the user or reader of its multifarious nature within the practical constraints of print publication, has been described by Robins (2004) as *disjunctive editing*.<sup>74</sup>

While usually associated with the recording of multiple textual stages contained in a single document, the *genetic* approach can also be used to edit texts whose genetic stages are recorded in more than one document, producing editions known as *synoptic* editions, a term borrowed from classical scholarship (Kline and Holbrook Perdue 2008: 189-90).<sup>75</sup> In the German tradition of *Editionswissenschaft*, this application of the genetic method to different documentary versions of the work is known as the *historical critical* approach. Editors of this school are not interested in authorial intentions, but rather in “each historical, physical iteration of the text” as “a social / linguistic event, with its confluence of forces and contexts of signification” (Modiano, Searle and Shillingsburg 2004: xii), in the vein of Jerome McGann’s social textual criticism. Its focus on the establishment of the chronological history of the text’s development means that this approach requires manuscript versions that represent consecutive revisions of the work, either by the same author or by a series of copyists, which excludes it as an editorial method for most medieval texts.<sup>76</sup>

<sup>73</sup> For example Donald Reiman and Jerome McGann, among others, have advocated what Reiman calls “‘versioning’ rather than ‘editing’: giving the reading public equally convenient access to more than one version of a text rather than a single clear text from which the various prior versions would need to be laboriously reconstructed from textual notes” (Kline and Holbrook Perdue 2008: 188).

<sup>74</sup> Another form of disjunctive edition suggested by Robins (2004) involves the juxtaposition of a single-text edition of the ‘best-text’ version with a similar edition of what could be called the ‘worst text’ version, or an eclectic text with one or more diplomatic ones, with the aim of providing the reader with “some sense of the range of discrepancies that inhere within the textual tradition” and to (148). As an actual example of a disjunctive edition, Robins (2004: 153) presents Rychner’s 1958 edition of *Lai de Lanval*, which juxtaposes a stemmatic edition at the top of the page against single-text diplomatic editions of four manuscript versions.

<sup>75</sup> The method and the term were first applied to an edition of a modern work by the editors of Joyce’s *Ulysses* (Joyce 1984), and it has since been applied not only to literary works, but also to historical documents such as the US Constitution and other legislative bills.

<sup>76</sup> This also includes the *Potage Dyvers* recipe collection, whose surviving versions bear an indeterminate relationship to each other and most likely represent a multilinear documentary tradition.

### Parallel-text editing

Unlike the other representations of multiple source documents described above, which all involve compressing the source documents into a smaller number of editorial texts, the *parallel text edition* is a more straightforward solution to the problem, based on the presentation of a full documentary rendition of each witness with no truncation or conflation. Not surprisingly, parallel text editions have been subjected to much the same kind of criticism as documentary editions in general, but to an even greater degree.<sup>77</sup> As with documentary editions in general, the objection most commonly levelled against parallel text editions is a moral one, the editor being “seen as abrogating certain functions and leaving the reader to make decisions and choices that should have been made for him/her” (Fellows 1998: 22).<sup>78</sup> In addition to this ideologically based criticism, parallel text editions have also been criticized for more practical problems, such as being confusing and “clumsy in use”, resulting in their “elaborate textual notes being ignored as the reader struggles to keep his place in a multiple of texts of the work” (Allen 1987: 13). Fortunately, these practical problems, along with the problem of being expensive to print, can be significantly mitigated by the technical means afforded by the digital edition, as will be described in chapters 4 and 5, and demonstrated by the sample editorial outputs included in this thesis.

Although parallel text editions have become a more acceptable solution over the last two decades, both because of the New Philological emphasis on the materiality and variability of medieval manuscripts, and the increasing awareness of their significance as reflections of linguistic variation, the text-critical role of the interventionist editor still looms large. For example Fellows (1998), while advocating parallel-text editions, still sees the role of the editor as essentially interventionist, observing that it is not enough “for the editor to take a strictly noninterventionist stance and simply present diplomatic transcriptions of manuscript texts” (23).<sup>79</sup>

Thus, while advocating the parallel-text edition as the preferred—or even the only sensible—option for certain types of texts, both Fellows (1998) and Pearsall (1994) point out that the parallel-text edition should not be seen as “a panacea for all editorial ills, but simply [...] one among a number of editorial approaches, whose suitability must be assessed on the merits of the individual textual case” (Fellows 1998: 23). These merits have usually been evaluated in terms of the type of source text, while the *intended use* of the edition is rarely mentioned.<sup>80</sup> Generally,

<sup>77</sup> Since parallel-text editing is applicable to texts surviving in multiple copies and are thus in direct competition with critical editions, they have been seen by critical editors as an even ‘greater evil’.

<sup>78</sup> Before the 1990s, as Embree and Urquhart pointed out in their 1987 article, parallel-text editions were not very numerous and were often considered to be somehow inferior in scholarly quality (49). For example, George Kane and E. Talbot Donaldson, the paragons of eclectic editing characterized any edition “which stops short of a fully reconstructed original” as “poor-spirited and slothful undertaking” (Kane and Donaldson 1975: 129). This criticism was still being felt by scholars defending parallel text editions as a valid scholarly endeavour twenty years later, as both Pearsall (1994) and (Fellows 1998: 22) still feel that “[s]evere textual critics” (Pearsall 1994: 117) see parallel text editions as ‘noncritical’ and “pusillanimous, lacking in judgment, or motivationally dispossessed” (Fellows 1998: 22).

<sup>79</sup> Quite tellingly, she justifies this need for intervention by arguing that not “even a holograph manuscript will by definition represent exactly what the author intended at all points” (23), postulating authorial intention as the principal guiding principle of editing.

<sup>80</sup> From the point of view of editing for the purposes of historical linguistics, this seems like a rather



Moffat (1998) sees parallel text editing as justified in cases where—due to damage or some other reason—there exist texts that can be judged to represent different “versions of the same work” which are “not reducible to a single critical text” (Moffat 1998: 41), which is—as I will argue in chapter 9 and demonstrate in chapter 13—exactly the case with the *Potage Dyvers* collection.

One genre of Middle English texts with which parallel-text editing “has found increasing favor in the past thirty-odd years” are Middle English romances, which have traditionally been considered ‘uneditable’ by critical methods because of the large amount of variation between different manuscript versions caused by the continuous adaptation and augmentation of the texts by copyists (Fellows 1998: 21) and their “relatively unremarkable and non-individualistic mode of composition” (Allen 1987: 13). For example Embree and Urquhart (1987) adopt a parallel text edition as the vehicle for the presentation of the Middle English poem *The Simonie* because the text does not seem to fit the “conventional assumptions about medieval composition and transmission” (49). Their decision was based first of all on the recognition of its three separate versions “as quite distinct though closely related texts which need to be understood in relation to one another before they can be very seriously studied” (58), and subsequently on the fact that none of the other options traditionally available would do justice to this observation, as

to publish three separate editions would simply make the task of comparison physically awkward; to ‘restore’ the archetype would be to discard dozens of stanzas and scores of lines of Middle English verse; to combine all the stanzas of all three versions into a single ‘composite’ edition would be to edit a text that never was.

(Embree and Urquhart 1987: 58)

The same situation can be seen to obtain for the *Potage Dyvers*, as well as for many other vernacular practical and scientific texts, which have also been proposed as candidates for parallel text editing. However, Grund (2006) has argued that while this may be a practical solution for a short text, it is simply not practicable for an extensive text with multiple witnesses—like the *Mirror of Lights* edited by Grund (or the *Potage Dyvers* recipe collection)—as it would “mean serious layout problems and would probably make a publisher balk” (118). While this is undoubtedly true in the case of printed editions, digital technology allows us to overcome these problems, which—as will be argued in section 4.2—are not inherent to the parallel text edition, but rather to the printed medium.

More recently, Roland (2004) shifts the argumentation for parallel text editions away from their suitability for certain kinds of texts and more towards their suitability for certain *research purposes*, which is also the main argument of the present edition. She sees parallel text editions more as a *supplement* rather than a replacement for traditional ‘author-centric’ critical editions, useful for providing “additional avenues of textual inquiry that study the uses, uncertainties, and appropriations of a given work” (38), and disrupting “singular readings of a work” (49) in the teaching of medieval literature.<sup>81</sup> To these fields of application, I would

---

alarming oversight, considering that the requirements posed by this use (which will be discussed in detail in section 4.3) place quite stringent restrictions on the editorial methods used for this kind of an edition.

<sup>81</sup> In the case of medieval culinary recipes this disruptive function is welcome for dispelling the il-

obviously add historical linguistics, as the ability of diplomatic parallel text editions to analytically represent and juxtapose multiple textual variants of the same work, each with different contextual parameters<sup>82</sup> makes them well-suited to answer the demands of variationist historical linguistics, as will be argued in section 4.3.

Thus in the case of the present edition, both the nature of the material and its intended use speak for an editorial approach based on the presentation of parallel texts. The same conclusion was also drawn by Scully (1988) in his edition of the French recipe collection known as *Viandier de Taillevent*. In addition to acknowledging the impossibility of consolidating the different versions into a critical text without doing undue violence to them (Scully 1988: 9–10), he also explicitly names the study of manuscript relations as one of the intended uses of his edition (Scully 1988: 11). In a printed edition, this facilitation of contrastive analysis incurs some drawbacks in the visual presentation of the material, as the alignment of the contents of the different versions necessarily creates “the side effect of a fragmentary editorial form” (Roland 2004: 51).<sup>83</sup> However, the digital presentation medium will allow one to circumvent this problem at least to some degree by encoding into the edition information about both the original layout of individual manuscript versions and the parallel links between the different manuscript versions and using them to generate several alternative presentations between which the user can then move at will.

However, as was already pointed out, parallel-text editions are not the proverbial silver bullet, but rather a particular means of investigating a work surviving in multiple versions. Despite the benefits they can offer for the *analytical examination* of textual variation and the transmission process of a work, they are similar to eclectic or critical editions in the sense that as a whole, they also create “a text that never existed as represented” (51). Therefore, this edition takes the viewpoint, following Roland (2004), that parallel text editions are not “merely pragmatic responses to corrupted texts” (38), but rather a conscious, theoretically driven choice based on a specific view of medieval textuality and the nature of the source documents, described in chapter 2, and the intended use of the edition as a historical corpus linguistic resource.

### 3.3 Conclusion

For most of the 20<sup>th</sup> century, literary scholarship, particularly in English, privileged the author and used bibliographical and critical processes in order to reach a putative authorial text hidden or corrupted by subsequent error. Historians, by contrast, have generally preferred diplomatic editions (i.e. a text faithfully transcribed from its appearance in a particular document) or a type- or photographic facsimile of a particular document. (Walsh 2010b: 157)

---

lusion held by many amateur enthusiasts of medieval cookery—as a result of critical editions and modernized versions—that there ever existed set, canonical versions of individual recipes or dishes.

<sup>82</sup> E.g. different scribal idiolect, time and place of production, and possibly also target audience.

<sup>83</sup> This is also demonstrated by Scully’s edition, which completely sacrifices the original layout and even order of the manuscript versions for the parallel presentation of the texts.

The editing of historical utilitarian texts has suffered from falling in between the two major editorial traditions: *critical editing*, which has focused on literary works, and *documentary editing*, which has focused on unique historical documents. This means that there are no established ways of editing utilitarian texts, which do not fit the assumptions of either of these editorial approaches. On the one hand, they lack a strong authorial function and artistic aspirations and were thus often adapted quite freely by their copyists, making them ill-suited for critical methods of editing. On the other hand, they often exist in multiple, often quite significantly differing copies, making them equally ill-suited for traditional documentary editing. This means that they require a new editorial approach that takes into account their variability and rhizomatic textual transmission, representing both the unique features of each manuscript version and the relationships between them.

While critical editing is usually done from the viewpoint of the literary critic and documentary editing from the viewpoint of a historian, editions prepared from a distinctly *linguistic* point of view have not been very numerous.<sup>84</sup> For this reason, what is needed is an editorial method that takes into account the requirements of historical linguistics and allows us to accurately model textual objects as material witnesses of past linguistic acts. While such a method is by definition more *documentary* than *critical*, it will also need to be able to present multiple versions of the same work in a way that allows for the detailed investigation of the variation typical to medieval utilitarian works.

As has been repeatedly argued in this chapter, the editorial principles and practices followed by an editor should be determined not only by the social, historical and cultural context in which the edited documents were produced (Minnis and Brewer 1992: ix), but also by the particular purpose for which they are edited (McGann 1991: 47). In order to be able to question established editorial practices and to develop more effective ones, we thus need to be aware not only of the textual and historical characteristics of our source material, but also of the needs of our target audience and the practical means available to us in trying to meet them:

The philologist as producer should become familiar with the demands that current and future users make or will make on the product, and the philologist as user should with the producer's help recognize the production requirements and efficiency, the possibilities and limitations of the product.  
(Zeller 1995a: 18)

Fortunately, since the methodologies of both critical and documentary editing have developed in the context of print publication, many of the traditional practices that are problematic from the point of view of either medieval manuscripts or linguistic research are simply compromises necessary for the representation of manuscript content in a printed format. This means that for editions prepared in a digital environment, many of these practices can be disposed of as unnecessary, providing the editor with more freedom in accommodating both the source material and audience of the edition.

While this chapter has focused on the critical examination of the established principles and practices of critical and documentary editing from the point of view

<sup>84</sup> Nevertheless, it is specifically historical linguists that have in recent years expressed interest in the nonliterary, utilitarian and everyday writing of the medieval and early modern periods.

of digital editing of medieval utilitarian texts for the use of historical linguistics, the next chapter will take a more constructive approach and systematically outline the requirements of linguistic editing and the ways in which the digital medium will allow us to answer them. On the level of editorial ideology, this will involve a move from seeing the *edition* as a scholarly product or an endpoint, to seeing it as a scholarly tool, a starting point for scholarly research. This entails a radical shift in the editorial role, often reviled by critical editors as ‘un-editing’, where editors cease to be arbiters between truth and corruption, good and bad readings, and instead take it as their aim to “make as much about a text visible to a wide an audience as possible, rather than silencing opposing views or establishing one definitive text over all others” (Smith 2004: 310):

In other words, integrity in textual matters lies in *access to the evidence*, such that someone that follows us can see precisely how we worked, what we considered, so as either to ratify the evidence or object and think it through again. (Searle 2004: 14)

## Chapter 4

# Digital editions for historical linguistics

A medieval text is a moveable and organic object. Not only does it change in every manuscript witness, but it interacts on the page with illustrations, glosses, historiated capitals and lemmata. In some cases the medieval page is a hive of activity, full of visual stimuli with the text itself off-centre to make space for marginal illustrations and glosses that parody or interpret text, and lemmata that guide the reader who attempts to assimilate and synthesise what can only be called a multidimensional visual experience. For this reason modern technology, in particular hypermedia, has been able to restore to the reader the vitally important *context* of the medieval text lost since the invention of printing. (Caie 2000: 30-31)

As was concluded in the previous chapter, the appropriate editorial choices are contingent not only on the type of material edited, but also on the *purpose* for which the edition is intended and the *medium* in which it is realized. Until recently, virtually all theoretical discussion on editing texts surviving in multiple documents was focused on the practices of textual criticism and the production of critical editions. It was the development *digital editions* in the 1990s that brought about a serious challenge to the hegemony of critical editing on the level of both practice and theory. And it is only in the last 20 years that the editorial community has started to seriously account for the possibility that many of its established truths may in fact be mere side-effects of the printed medium and to formulate theoretical approaches specifically for the digital medium.

While this edition is decidedly *documentary* in its orientation, the fact that it is intended for *linguistic* research and is implemented in the *digital* medium provide this documentary orientation quite different implications from those of the printed historical documentary edition. While historians are primarily interested in the textual *content* of the original textual object, historical linguists are interested primarily—although not exclusively—in its textual *form*. And while a printed

edition is an inherently static object, having as one of its primary functions the establishment of a stable text, a digital edition, as will be argued here, is by its nature a *dynamic* model, a flexible tool for the exploration of the textual object. The different requirements posed to the edition by its use, and the different possibilities afforded to it by its medium thus mean that the very image of the 'ideal edition' in this context is bound to differ from the traditional.

Since the development of the first machine-readable *language corpora* in the 1970s, the emphasis of linguistic study has constantly shifted towards the empirical observation and analysis of actual linguistic performance. The adoption of corpus linguistic methodology also by English historical linguists created a need for computer-readable renditions of historical texts. Initially, this need was answered principally by digitising traditionally edited versions of historical texts, which provided the linguist with a stable established text for analysis. However, over the last decade or so, there has been a growing realization that not only is the whole idea of an established text an anachronistic fiction in the context of inherently variant and unstable manuscript textuality, but it is actually the same phenomenon of *variation* that lies at the core of historical linguistic change. The interest in variation has also raised to the foreground the concept of context, as linguistic variation has been found to be conditioned not only by language-internal factors, but also by the cultural, social and physical context of the communicative act. This has resulted in an increased interest in the role of the physical document itself as a communicative device that not only forms the immediate physical context of the communication it mediates, but also encodes various aspects of the cultural and social context in which that communicative act took place. These developments have resulted in a demand for research materials that make use of the capabilities of the digital medium in order to build detailed models of historical documents instead of merely providing an editorially established account of their abstract textual content.

One of the principal aims of this thesis and the present edition is to provide a model for producing such resources, and this chapter is intended to outline the general requirements of a *corpus linguistic edition* and the ways in which the digital medium allows us to fulfil these requirements. As McDermott and Walsh (1991: 43) have argued, the editorial viewpoint on the text overrides even the characteristics of the edited material. Just as a work originally conceived as an instrumental or religious text can later be edited as a literary text, *any* kind of text can also be edited as a linguistic object, a fossilized record of a linguistic act that took place at a specified point in time and in a specific context, resulting in a *linguistic edition*. The first section of this chapter will examine the interface between philology and linguistics and introduce the theoretical background of *variationist historical corpus linguistics*, which constitute the primary intended application of the present edition and the editorial method outlined in this thesis. The second section will introduce the concept of a *scholarly digital edition*, and present an overview of the benefits that it can provide in comparison with its more traditional printed counterpart. The final part of this chapter will then proceed to discuss the specific requirements posed by historical corpus linguistics for this kind of a digital scholarly edition.

## 4.1 Philology and linguistics

*Philology* as a term has come to be associated with the gathering and recording of linguistic data while *linguistics* is associated with rigorous theoretical analysis of this data. However, this distinction is not very useful, since the two “are richly intertwined in English historical linguistics” (Curzan and Emmons 2004: 5). As English linguistics has over the last few decades moved increasingly towards the empirical observation of actual language use through linguistic corpora containing authentic samples of language use in digital form, historical linguists interested in the earlier stages of English have been at a disadvantage. Not only do they have to content themselves exclusively with written material, but the compilation of historical corpora involves challenges that compilers of modern corpora never need to encounter, such as “multiple versions of manuscripts, scribal errors or poor physical document quality” (Dollinger 2004: 5). To solve these problems, which lay squarely within the purview of traditional philology and require skills outside those of corpus linguistics, many corpus compilers have ‘outsourced’ the philological aspect of the process to scholarly editors and compiled their corpora from printed editions of historical texts, which “have generally not been produced with linguistic study in mind, and may not always be reliable” (Kytö, Grund and Walker 2007).

While the need for more linguistically oriented editions that “aim at reproducing the original manuscripts more faithfully than critical or eclectic editions do” (Kytö, Grund and Walker 2007) has been widely acknowledged among historical linguists in recent years, the lack of integration and communication between the disciplines of (digital) manuscript editing and (historical) corpus linguistics has resulted in unnecessary duplication of effort and in the production of digital editions that either omit linguistically significant information or restrict access to their source data, excluding historical linguists from among their users (Honkajärvi, Kaislaniemi and Marttila 2009: 452). In this respect the present edition joins the call, articulated for example by Meurman-Solin (2001) and Dollinger (2004), for a closer connection between philology and historical corpus linguistics and the promotion of “philological computing” (Meurman-Solin 2001: 18–24) in order to avoid the loss of linguistic precision entailed by the outsourcing of the philological part of research to editors with their own, often decidedly non-linguistic aims.

### 4.1.1 Theoretical background: New Philology

One can learn a great deal from the variations, additions, and glosses that accompany the text in its various presentation, though it is just these things that a Lachmannian edition would ignore or relegate to the obscurity of an *apparatus criticus*. Instead, the text should be presented in a form that conveys its meanings in its different medieval and early modern settings. (Diehl 2004: 59)

The theoretical underpinnings of the present edition and the following discussion of digital linguistic editing lie in the philological school known as *New*

*Philology* that emphasizes understanding the written sources of a culture from the perspective of their original context and ‘textual culture’. Originally developed in the field of ethnohistory in the 1970s, its first application to medieval manuscript studies has been considered to be the 1989 essay of Bernard Cerquiglini, titled “Eloge de la variante”, which emphasized the concrete and factual materiality of texts (Bodard and Garcés 2009: 95-6) and the inherently variant nature of medieval writing, as the oft-cited passage<sup>1</sup> from Cerquiglini points out:

La recriture incessante à laquelle est soumise la textualité médiévale, l’appropriation joyeuse dont elle est l’objet, nous invitent à faire une hypothèse forte: la variante n’est jamais ponctuelle. C’est l’énoncé lui-même que travaille, comme une pâte, l’activité paraphrastique; ce n’est pas par le mot qu’il convient de saisir cette variance, mais pour le moins au niveau de la phrase, voire même au sein de l’énoncé complet, du segment de discours.<sup>2</sup> (Cerquiglini 1989: 111)

This means that different manuscript versions of the same work cannot be reconciled with each other on the linguistic or even textual level, but only on the abstract discourse level of the work. In addition to embracing the *variance* of medieval writing, New Philology has also broadened the philological concern beyond just text and language, taking into account the whole of the medieval “manuscript culture”, including also the visual and physical properties of manuscripts like layout, images, rubrication, etc. (Nichols 1986: 7). The New Philologists can thus be seen as going even further than McGann in not only arguing “for a renewed interest in the material contexts of medieval literature, what they often call the manuscript culture”, but also entirely dispensing with the text-critical emphasis on the author (Moffat 1998: 47). As Shillingsburg (1991: 41) has pointed out, this emphasis on the multiplicity and indeterminacy of texts was not received with unanimous acceptance, as all of the pre-1990s editions that attempted to “emphasize alternative texts, or multiple texts, or indeterminate texts”, such as Hans Walter Gabler’s edition of *Ulysses*, Michael Warren’s edition of *King Lear*, and Shillingsburg’s own Thackeray edition were still considered “controversial” in 1991.<sup>3</sup>

### Criticism of New Philology

Scholars have objected to the New Philological agenda on various grounds. Some, like Derek Pearsall, have agreed with the New Philological emphasis of textual variation and the materiality of texts as a requirement for understanding medieval texts (Hanna 1992: 125), but wanted “to have nothing to do with” their deconstructivist aim of getting entirely rid of authorial intention: for Pearsall “authors and

<sup>1</sup> Both Nichols (1986) and Fleischman (1986) quote it in their articles published in the special issue of *Speculum* dedicated to New Philology (vol. 65).

<sup>2</sup> Translated in Cerquiglini (1999: 78) as: “The endless rewriting to which medieval textuality is subjected, the joyful appropriation of which it is the object, invites us to make a powerful hypothesis: the variant is never punctual. Paraphrastic activity works on the utterance itself, like dough; variance is not to be grasped through the word; this must be done, rather, at least at the level of the sentence if not, indeed at the very heart of the complete utterance, of the segment of discourse.”

<sup>3</sup> McGillivray (1994: 181) has suggested that the New Philological project and Cerquiglini’s *Éloge de la variante* caused “considerable furor in the English-speaking world” not because of the outlandishness of its ideas, but precisely because it openly articulates “an anxiety about the authenticity of the critical edition” that was already familiar to the editorial community.



their intentions, conscious or unconscious, are of the very greatest significance”, and he cannot imagine “how any practice of textual criticism could ever detach itself from the attempt to determine an author’s intentions” (Pearsall 1994: 125). On the other hand, others like Tim William Machan, who see medieval textuality as emphasising the *res* over the *verba*, consider the New Philologists’ concentration on “the material manifestations of the manuscript culture” as misguided (Moffat 1998: 47).<sup>4</sup>

Perhaps the most total rejection of the New Philological view has come from Varvaro (1999), who not only dismisses Cerquiglini as “too misinformed and/or unconstrained to be credible” (57) but takes a quite smug and arrogant attitude towards the whole idea of New Philology, claiming that Italian textual scholars are “more wary of Foucault and Derrida, and unlike the Americans [...] have not been blinded by deconstructionism” (57). On a more practical level, Varvaro (1999) also expresses a fear that the kind of “‘Selbstbedienung’ (self-service) of the text from the computer screen” advocated by Cerquiglini “will pave the way for the triumph of incompetence and for the creation, however ephemeral, of innumerable inauthentic texts” (Varvaro 1999: 57).<sup>5</sup> Fortunately, times seem to have changed somewhat, since as (Echard and Partridge 2004a: xii) point out, many younger scholars no longer see New Philology as a departure from existing tradition but rather a “formative influence” or even an established tradition itself.

### Implications for editing medieval texts

If we accept the multiple forms in which our artifacts have been transmitted, we may recognize that medieval culture did not simply live with diversity, it cultivated it. The ‘new’ philology of the last decade or more reminds us that, as medievalists, we need to embrace the consequences of that diversity, not simply to live with it, but to situate it squarely within our methodology. (Nichols 1986: 8-9)

Cerquiglini (1989) has theorized medieval texts from the point of view of deconstructive theory and argued “that medieval writing forms the ultimate post-modern *texte*: the individual work automatically dissolves into the plurality of all its variants, and its meaning resides precisely in such *mouvance* or *variance*” (Hanna 1992: 121). Cerquiglini’s central thesis is that medieval writing does not merely *contain* more or less authorized variants on the level of individual words, but being perpetually unfinished, copied by hand and constantly manipulated, it *consists* of endless variance resulting from constant “intervention, annotation, and commentary” (Cerquiglini 1999: 34).

<sup>4</sup> Even within the ranks of scholars openly supporting the agenda, there seem to be some misgivings; for example Wenzel (1986), while embracing the New Philological emphasis of original documents and the variability of the text, is clearly not willing to let go of the image of the heroic philologist establishing the one true text, quite stoutly declaring Kane and Donaldson’s work “a monument of philology” (13).

<sup>5</sup> He is somewhat unclear as to what he means by ‘inauthentic’, but his general argumentation would suggest that he is referring to texts that fail to attempt the reconstruction of an authorial original.

Since variance is the primary characteristic of medieval textual objects, Cerquiglini (1999: 78) argues that editions must be designed in a way that gives priority to this variance and allows us to analyse it effectively. According to the New Philological view, the appropriate editorial reaction to manuscript variation is thus not a quest for the single variant closest to the elusive *Urtext*, but rather a comparative analysis of the variation and its contribution to the meaning of the text (Fleischman 1986: 25). Upon encountering several variant versions of a passage in a medieval text, one “need not attempt to discover which is closer to the ‘original’ (a philologist’s reflex), or even which is the older (grammatical reflex); one must assume their equivalence and grasp medieval language as it swings back and forth between the two” (Cerquiglini 1999: 75). Individual textual witnesses are thus seen as valuable in their own right, as their individual textual and paratextual qualities tell the scholar something about the cultural and situational context in which they were produced (Carlquist 2004: 112).

This focus on the different versions of individual texts and on the proliferation of these texts’ intrinsic variants has inevitably resulted in “a renewed interest, among the practitioners of New Philology, in manuscripts and their material status” (Gumbrecht 1998: 248). This movement of focus from the *work* towards the *document* has resulted in what Gumbrecht calls “the heuristic presupposition of a weak editor-subject and a weak author-subject” (248). According to him, the ‘weakness’ of editorial and author roles refers “to a philological practice where, on the author-role level, the process of transmission receives more attention than individual authors, and where, on the editor-level, the accurate rendering of the texts constitutes a more important task than their manipulation and modification” (248). Gumbrecht sees the influence of the “traditional philological principle of text-based evidence” (247) to have been similar in philology to the influence of strong truth-concepts in philosophy, namely that it has led philologists on a pursuit of the single authoritative text and of “‘right’ answers and ‘correct’ solutions”. As a counter to this traditional text-critical view, this thesis and the present edition takes what Gumbrecht calls a “linguistico-pragmatic approach to text-editing”, abandoning “the idea of the one ‘correct’ edition as the ultimate telos” (247) and replacing it with a multiplicity of possible editorial roles, producing different kinds of editions for different purposes and different communities of readers. This means that instead of seeing different types of editions—such as a documentary, critical or what Gumbrecht calls ‘neo-philological’ edition—as ‘better’ or ‘worse’, they should be seen as more or less suited for a given purpose.

As was stated in chapter 1, the purpose of the present edition is to serve as research material for historical variationist corpus linguistics, which means that its representation—or *modelling* (see section 5.4)—of the original document has been tailored primarily for the needs of this approach. Since the *variationist* approach to historical linguistics is—rather unsurprisingly—interested in the variation between different realizations of a linguistic function, the edition will also serve as a practical demonstration of a ‘neo-philological’ edition that tries to represent the inherent variability and instability of a medieval work on both textual and documentary levels in a format that allows us to flexibly visualize and analyse it.

### 4.1.2 Historical linguistics

Historical linguistics is sometimes equated with *diachronic linguistics*, referring to the study of how language changes in time (see e.g. Campbell 2004: 1). Concerns that are commonly seen to fall within the purview of diachronic historical linguistics include 1) the description of diachronic changes occurring in a particular language, 2) the reconstruction of the pre-history of languages and their grouping into 'language families' (comparative linguistics), 3) the development of general theories about the mechanisms and reasons of language change, and 4) the history of individual words, (i.e. etymology) Campbell 2004: 4–6. However, this thesis will adopt a slightly wider definition, used for example by Janda and Joseph (2003: 85–6), which blurs the distinction between synchronic and diachronic linguistics and considers historical linguistics to cover both diachronic linguistics and the synchronic linguistics focused on reconstructing the linguistic system of a specific historical point in time. The reasons for this are twofold. First of all, the establishment of the synchronic state of linguistic systems at different points in time is a necessary prerequisite of diachronic linguistics, and secondly, as Lass (1997: 9–16) has argued, the distinction is largely artificial and not very useful, since the synchronic structure of a language is always the result of and a participant in a historical process and cannot be understood without taking the diachronic viewpoint into account. However, linguistic change is here understood as highly contextual. In other words, the processes of linguistic change are conditioned not only by the physiology and cognition of humans and the internal structure of the language, but also by social and cultural factors (Campbell 2004: 316–7; Rissanen 2008: 56).<sup>6</sup>

Language in the sense of Saussure's *langue* is in this thesis seen not as a homogenous system, as argued by Saussure and generative linguists, but as a theoretical construct formulated by the linguist as a description of the regularities observed in the heterogeneous system made up of the individual linguistic practices of language users. This, as Weinreich, Labov and Herzog (1968) have observed, does not mean that it is not structured, but rather that in any real community, there is always some structured heterogeneity in the language used by its members (101). Furthermore, the linguistic competence of an individual is also seen as an orderly and heterogeneous system containing alternative linguistic forms for accomplishing similar tasks, and in which "the choice between linguistic alternants carries out social and stylistic functions" and which changes "with accompanying changes in social structure" (Weinreich, Labov and Herzog 1968: 161–2). Thus the only way of investigating the language as a system is to investigate its realization in individual utterances made by users of the language. Because of this emphasis on individual linguistic utterances, a linguistically oriented edition is not concerned with the recreation of the original reading (or writing) experience but rather with their *analytical examination*, which makes as direct access to

<sup>6</sup> This means that the method advocated for example by Janda and Joseph (2003) for the study of historical linguistics, namely observing linguistic changes occurring in the present and using them as analogies to changes in the more distant past is a dangerous one, as it ignores the role of the extralinguistic social context in which these changes took place and which are vastly different for example between the present day and late medieval England. Therefore we must always be wary of generalising observations across different time periods and consider them as tentative hypotheses until we can test their probability against historical evidence.

the relevant surviving evidence—i.e. the original document—the first priority.

The branch of historical linguistics to whose needs the present edition has been designed is thus an empirical approach to the study of language change that could be characterized as *historical variationist corpus linguistics*. In order to provide some context for understanding what it means to produce a digital edition for the needs of historical corpus linguistics, this subsection will briefly outline the central features of both corpus linguistics and variationist linguistics.

### Corpus linguistics

The introduction of corpora has had a revolutionary effect on language studies in the last few decades. This is particularly true of historical linguistics, which has to rely on written sources only; introspection and native-speaker competence cannot be relied on in the study of the language of previous centuries and millennia. We could even suggest that in the present world the creation of corpora has been a matter of life or death for the future of evidence-based historical linguistics, at least in the study of extensively spoken living languages. (Rissanen 2008: 53)

Corpus linguistics is a method of studying the use of language through the quantitative and qualitative analysis of samples of naturally occurring language known as a *linguistic corpus*, defined as “a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research” Sinclair (2005), or “a finite electronic collection of texts or parts of texts by various authors which is based on well-defined and linguistically relevant sampling criteria and aims for some degree of representativeness” (Claridge 2008: 242). Its aim, unlike more structurally oriented forms of linguistics, is to empirically investigate “how speakers and writers exploit the resources of their language” (Biber, Conrad and Reppen 1998: 1). This investigation of language use has two central research goals: first of all, assessing the extent to which a certain feature is found, and second, analysing the contextual factors that influence variability between alternative realizations of a linguistic function (Biber, Conrad and Reppen 1998: 3). Thus the methodology of corpus linguistics is intimately tied with the *variationist* approach, as it focuses not only on the different ways in which language is used, but on the various factors—genre, register, textual context—that constrain and influence its use.

Corpus linguistics has often been associated with purely quantitative studies and in the early stages of computerized corpus linguistics these were indeed in the majority, largely due to the newfound ease with which calculations could be performed on machine-readable texts (Johansson 2008: 48). However, nowadays it is almost universally recognized that “corpus-based analyses must go beyond simple counts of linguistic features” and provide “qualitative, functional interpretations

of quantitative patterns" (Biber, Conrad and Reppen 1998: 5), or as Rissanen (e.g. 2008: 66) has often reiterated: "Research begins where counting ends".<sup>7</sup>

Although the basic idea of corpus linguistics predates the use of computers, it was the access to "machine-readable texts which could be stored, transported and analysed electronically" (Johansson 2008: 33) that caused a breakthrough in the 1970s. The term *corpus linguistics* itself, referring specifically to 'computer corpus linguistics', was introduced after electronic corpora had already been used for some time (34).<sup>8</sup> Although the first computerized English language corpus, the influential *Brown Corpus* (see Francis and Kučera 1979), was compiled already in the 1960s, it was after 1970 that the compilation and use of computer corpora started to gain momentum (Johansson 2008: 34).<sup>9</sup> As computers became more powerful, cheaper and more user-friendly in the 1980s, linguists became less dependent upon outside computational expertise, and the quantity and variety of digital texts exploded. While the potential of computers to provide "an unprecedented way of studying language in use" had been widely recognized by the end of the 1980s, there persisted a tension between corpus linguists and armchair linguists, especially early generative linguists (33). In terms of historical linguistics, the most important undertaking in corpus compilation was the compilation of the diachronic part of the Helsinki Corpus of English Texts in the late 1980s under the direction of Matti Rissanen (Johansson 2008: 42-3), which can be seen as one of the projects that led to the establishment of corpus linguistics as a part of the mainstream of linguistics in the 1990s (49).<sup>10</sup> While modern language corpora have benefited enormously from the increasing emergence of computer-readable texts from the 1980s as by-products of computer typesetting (Johansson 2008: 40), historical corpora based on manuscript or early printed texts have not experienced a corresponding benefit and still need to be keyboarded manually.<sup>11</sup>

Although the term *corpus linguistics* is often used as if it denoted a specific linguistic discipline, it was already in the 1980s recognized as "a tool that could be applied in virtually any branch of linguistics" (Johansson 2008: 34) rather than

<sup>7</sup> It is true, however, that in terms of the dichotomy between text as quantifiable data to be picked apart and counted on the one hand, and as indeterminate works of art to be appreciated and interpreted on the other (Flanders 2009), corpus linguists have traditionally been placed firmly in the former camp.

<sup>8</sup> Its first published use was in a collection of papers from the 'Conference on the Use of Computer Corpora in English Language Research' held in Nijmegen in 1983, which was titled *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research* (Aarts and Meijs 1984).

<sup>9</sup> In order to establish co-operation in the use of new technology for the analysis of texts, new organizations like the Association for Literary and Linguistic Computing and the Association for Computing in the Humanities were established in the 1970s (in 1973 and 1978, respectively) (Johansson 2008: 34). It is somewhat ironic that while intimately connected with the early phases of corpus linguistics, the community represented by these organizations later drifted apart from the corpus linguistic community and became known as 'Digital Humanities'. It has been only during the last 10 years that these communities have again begun to merge together.

<sup>10</sup> However, this might be a particularly European view of the state of the art, as Curzan and Palmer (2006) still point out that "corpus linguistics, at least in the United States, remains outside most definitions of 'mainstream' linguistics" (29).

<sup>11</sup> Various mass digitization ventures like the *Early English Books Online* (EEBO) and *Eighteenth-Century Collections Online* (ECCO) Text Creation Partnerships and Google Books are beginning to make digital texts available for compilers of historical corpora, although the quality of these texts and the conditions under which they are available still leave much to be desired.

a separate branch of linguistics (Rissanen 2008). As Biber, Conrad and Reppen (1998) have pointed out, the main benefit of corpora as a tool is that the empirical investigation of corpora “can shed new light on previously intractable research questions in linguistics” (ix) by providing a means of “handling large amounts of language and keeping track of many contextual factors at the same time” (3). This allows us to move beyond observing individual linguistic features and to focus on various kinds of “association patterns” or “systematic ways in which linguistic features are used in association with other linguistic and non-linguistic features” (5). Linguistic corpora also highlight the fact that these association patterns are not absolute but continuous: they do not describe what ‘always’ or ‘never’ happens, but rather what happens more or less frequently. Intuition or casual observation might tell us that certain patterns seem rarer or more common than others, but corpus methodology allows us to *quantify* the relationships between different variant patterns (8) and to overcome our human tendency to focus on the unusual and ignore the typical, providing a more objective view of language use than our linguistic intuition (3).<sup>12</sup>

As a research methodology corpus linguistics is especially well-suited to the New Philological viewpoint, as the use of corpora to collect and analyse material “helps us approach and appreciate the richness and variability of language and to understand how linguistic change is related to this variability, caused by both internal processes of change and language-external factors, socio-cultural, regional or genre-based” (Rissanen 2008: 54). In addition to these methodological advantages, corpora also provide scholars with several practical advantages. They can radically reduce the time spent on collecting evidence and leave more time for the analysis of the material, and significantly improve the verifiability and falsifiability of research results (Rissanen 2008: 64-5).<sup>13</sup>

Much of the criticism against corpus linguistics is actually focused on the shortcomings—perceived or real—of the corpora currently in use rather than the basic methodology itself, and should be seen as an encouragement for the improvement of corpora rather than a denouncement of the approach itself.<sup>14</sup> Despite its advantages, the corpus linguistic approach does pose several potential problems that scholars need to be aware of. Discussing historical corpus linguistics, Rissanen (1989, 2008: 65) has outlined the three major dangers facing the historical corpus linguist:

- 1) corpora only represent a part of linguistic reality (‘the God’s truth fallacy’);
- 2) for less-frequent linguistic phenomena, the figures of occurrence given by

<sup>12</sup> However, it should be kept in mind that while corpora can theoretically be seen as neutral resources that can be used for research representing a variety of theoretical standpoints, they are always designed and compiled within the context of a specific theoretical framework and the decisions made in their design and collection can “strongly constrain the kind of research that is carried out” (Hunston 2008: 166).

<sup>13</sup> Although it is important to remember that the effort of collecting research material has not been eliminated but rather relocated to the compilation of the historical corpora. The benefit of a well-designed and publicly available historical linguistic research resource like a corpus of course is, that it can be enjoyed by a large number of scholars and used for a large variety of purposes.

<sup>14</sup> An example of this kind of criticism, quoted in Curzan and Palmer (2006: 25), is the argument by Miller (1997: 252) that “the most fundamental problem [with such studies] involves methodology, reliance on modern technology, incomplete electronic samples of edited texts with no critical apparatus”.

corpora are too low for reliable conclusions ('the mystery of vanishing evidence'); and

- 3) drawing meaningful conclusions of corpus evidence requires a good command of the language form and an understanding of the literary, social and cultural background of the texts in the corpus ('the philologist's dilemma').

The close connection between historical corpus linguistics and philology also creates a problem with regard to traditional corpus linguistic methodology, originally developed in the context of modern corpora. This problem stems from the fact that while philology is essentially concerned with texts, the traditional practice in corpus compilation is to include not entire texts but *samples* or small, random chunks of text (Claridge 2008: 246). The problem is especially acute for New Philology which emphasizes the significance of the entire documentary context, which a sampled corpus completely obscures. This dilemma and the increasing interest in corpus-based historical genre and discourse studies have created pressure for compiling corpora consisting of longer extracts or even full texts. These kinds of corpora, especially if explicitly annotated for textual structure (see section 5.4), would allow us to study the relationships between linguistic features and textual structure. The same principle also applies to the annotation of documentary features like layout and typography; if they are explicitly annotated into the corpus, we can use corpus methods to analyse the association patterns between them and various kinds of linguistic and textual features.

Another central concept of traditional corpus linguistic methodology that is problematic for historical corpora is that of *representativeness*, referring to "the relationship between the corpus and the body of language it is being used to represent" (Hunston 2008: 160), which makes it possible to make statements not only about the corpus itself, but also about the body of language as a whole. While representativeness in terms of the entire language is extremely problematic even in the case of modern corpora, historical corpora are problematic even within a more restricted scope, since the vagaries of manuscript survival mean that it is usually impossible to estimate the extent of any target population with a reasonable degree of statistical validity (Claridge 2008: 247). Since "historical data have been accidentally preserved and are therefore not equally representative of all aspects of the language" (Milroy 1992: 45), different registers and varieties, as well as the language of different periods, may be unevenly represented in the surviving data, and we can have no reliable measure of these biases. The restricted amount of material available for practically all periods before the 19<sup>th</sup> century leaves the corpus compiler with two options: either include all surviving material, sacrificing balanced structure, or include principled samples of the surviving material, sacrificing corpus size (Claridge 2008: 245).<sup>15</sup>

In addition to these fundamental methodological problems, historical corpora also pose some practical challenges that users of modern corpora rarely need to deal with (Rissanen 2008: 65-6). First of all, the early stages of most languages

<sup>15</sup> This problem is especially acute in the case of spoken language, which can at best be represented very indirectly—or not at all—in historical corpora, despite playing "a decisive role" in the variation and change of languages (Rissanen 2008: 60). Even when recorded indirectly in written sources, historical records of spoken language—in addition to being unreliable as to the original phonology—are always to some degree "deprived of the social and situational contexts in which speech events occur" (Milroy 1992: 45).

exhibit significant amounts of orthographic instability, the same word being spelt in numerous ways, making it difficult to retrieve all occurrences of a word or a linguistic form. Second, most historical corpora only include a transcript of one manuscript witness of a text, obscuring any variation between different witnesses. Third, historical corpora are often based on modern printed editions rather than original manuscripts, which means that they preserve no record of the documentary context. Fortunately, all of these practical issues are more or less solvable, as will be shown in section 4.3 below.

### The variationist approach

The variationist approach to change sees linguistic variation and linguistic change as two faces of the same coin, two different aspects of the same phenomenon. All human speech-communities exhibit synchronic variation on a large scale, and language change across time is one outcome of this variation; conversely, linguistic variation is the inevitable synchronic face of long-term change. (Guy 2003: 370)

Whereas corpus linguistics is rather a methodology than a new ‘kind’ of linguistics, the *variationist* approach to language can legitimately be seen as a new branch of linguistics arising from the employment of computerized corpora (Rissanen 2008: 54). The concepts of variation and change through variation are not new, but were familiar already to nineteenth-century scholars who observed that languages have ‘different ways of saying the same thing’ and were interested in the “factors explaining the loss and emergence of forms” (Rissanen 2008: 55). However, its current popularity is intimately connected with linguistic corpora that not only offer “examples of the use of variants in various contexts”, but also make it possible to observe their relative frequencies (58).

The systematic study of variation as a means of explaining language change was brought into focus by a highly influential essay by Weinreich, Labov and Herzog (1968), and developed through the work of scholars like Samuels (1972), Romaine (1982) and others over the last fifty years (Rissanen 2008: 56). The approach has also been significantly influenced by M. A. K. Halliday’s discussions of language from a social perspective, following the views of Malinowski and Firth (Halliday 1973), and by the concept of *variant fields* or groupings of roughly synonymous variant expressions provided by a language for conveying the same meaning (Rissanen 2008: 55). The aim of the variationist approach is to analyse language by “describing the structure of the variant fields and comparing the characteristics of variants within the field, with special reference to the language-internal and external factors affecting the use of the variant” (55).

The variationist view of language “is built on the axiom that language is variable at all times” (Milroy 1992: 123) and that the potential for change is therefore always present. Therefore, as Guy (2003: 370) points out, echoing Weinreich, Labov and Herzog (1968: 159), linguistic features do not change abruptly and totally, but there is always a period during which the ‘old’ and ‘new’ forms exist simultaneously in a speech community. In diachronic variationist linguistics the



focus is on the changes that occur in the size and shape of these variant fields and the relative frequencies of different variants, which constitute language change. When language changes, some variants “become more common, unmarked or prototypical, while others may become rarer and restricted to certain contexts, genres or registers”, and occasionally some variants may even fall entirely out of use and new ones emerge to take their place (Rissanen 2008: 55).<sup>16</sup>

As mentioned, variationist linguists see the choice of the variant to be influenced by both extralinguistic and intralinguistic factors. Rissanen (2008: 56) groups the principal extralinguistic factors into 1) sociolinguistic (social status, education, relationship between discourse participants), 2) textual (genre, topic or purpose of text, discourse situation and medium), and 3) regional (including language contact), and describes the intralinguistic ones as “mainly related to basic patterns of change in meaning, such as metaphor or metonymy, grammaticalization, and the tendency to seek out new and more emphatic expressions to replace old and partially bleached ones” (58). In terms of this thesis and the present edition, historical variationist corpus linguistics constitutes the theoretical and methodological framework whose theoretical and practical requirements guide the design of the edition as a *model* of the original document, discussed in chapter 5.

## 4.2 Scholarly digital editions

The hypertext edition, the hypermedia edition, the multimedia edition, the computer edition, the digital edition, and the electronic edition are all synonymous labels for a concept without a definition.  
(Vanhoutte 2006: 161)

As was observed in chapter 3, the development of the conventions of scholarly editing over its entire history has been conditioned and constrained by the technology of print, resulting in a model of editing that is inextricably bound with the archetype of the printed book (Bøe, Jørgensen and Taugbøl 2004: 55). For editing documents that do not conform to that archetype, like medieval manuscripts, this model is thus “at best a limited, institutionally ratified starting point, and at worst an unnecessary discursive constraint” (Mussell and Paylor 2009: 139). Just as the availability of machine-readable text corpora has opened up new possibilities in the field of linguistics, the development of various digital encoding and processing technologies has opened up new possibilities for the representation and modelling of manuscript documents. While any published text can be called an edition and any text published in a digital format can consequently be called a ‘digital edition’, neither all texts published in print nor all electronic texts can properly be considered editions in the scholarly sense. The *Minimum Standards for Electronic Editions* of the Association for Documentary Editing (ADE) set the following requirements for a digital documentary edition: 1) rigorous attention to the text, 2) explanatory

<sup>16</sup> Not all variation leads to change, however, and it is possible for variant forms to exist in active alternation within a speech community for generations without any variant supplanting the others (Guy 2003: 371).

annotation and 3) an explanation of the editorial practices used on the texts (ADE Committee 2002).

Edward Vanhoutte, who has edited both literary and non-literary materials in digital media has defined an *electronic scholarly edition* (Vanhoutte 2006: 163) as an edition

- 1) that is the immediate result or some kind of spin-off product from textual scholarship;
- 2) that is intended for a specific audience and designed according to project-specific purposes;
- 3) that represents at least one version of the text of the work;
- 4) that has been processed from a platform-independent and nonproprietary basis, that is, it can both be stored for archival purposes and also be made available for further research (Open Source Policy);
- 5) whose creation is documented as part of the edition; and
- 6) whose editorial status is explicitly articulated.

This definition can be understood to incorporate the first and third requirements of the ADE definition (as items 1 and 5),<sup>17</sup> while making some interesting additions. The second item explicitly recognizes the impossibility of preparing a ‘general purpose’ edition and encourages editors to explicate the intended audience and uses of their edition. The fourth item can be seen as a political statement, but is actually a very reasonable criterion for a scholarly product: in order to be scholarly, it must be usable as research material, which in turn requires it to be ‘transparent’ to the user.<sup>18</sup>

Although some scholars like Sanders (1995: 129) have called the digital or electronic edition a new “genre” of edition, it does not really correspond to any commonly held definition of the term. Rather, the digital edition could more accurately be characterized as a presentation *medium*, or ‘modality’, parallel to the printed codex. Thus, as for example Dahlström (2009: 29) has pointed out, digital editions should not be considered a particular type or genre of edition, since they do not share a single predefined function or theoretical basis. This means that it is useless to talk about the differences between ‘digital editions’ and for example ‘eclectic editions’. Rather, the concept of a digital edition should be understood purely in opposition to a printed edition, each form allowing for a multitude of different editorial approaches: just as we have printed eclectic editions and diplomatic editions, we can also have digital eclectic editions and digital diplomatic editions.<sup>19</sup>

<sup>17</sup> The second requirement is not included in this list, and can be argued to be specific to the American tradition of documentary editing. The association of item 1 with the first requirement of the ADE definition is not obvious, but I interpret it as a requirement that the text of the edition be established following accepted scholarly practices.

<sup>18</sup> Although neither of these definitions fully encompasses the properties of a digital scholarly edition, as will be seen in chapter 5 and section 5.4, the present edition and the model of editing historical manuscript texts proposed in this thesis can be said to fulfil the requirements of both of these definitions.

<sup>19</sup> This means that in light of the conceptualization presented in chapter 2, digital editions could be seen as a distinct *register* (and consequently a distinct text type) by virtue of differences in the communication channel they adopt, i.e. the *mode* of their situational context (section 2.2. This interpretation would also agree with the observation of Tyrkkö (2011: 14-5) on hypertexts more generally, namely that they cannot really be considered to constitute a distinct *genre*, but they do constitute a distinct text type.

As McGillivray noted already in 1994, various editors and scholars (including e.g. Cerquiglini, Robinson, Deegan, Baker and himself) had raised the idea of a 'hypertext edition' in the late 1980s and early 1990s, seeing it as a superior medium for editing medieval texts (188).<sup>20</sup> The 1990s were a time of great optimism in terms of the new digital medium. For example Peter Baker had high hopes for computerized editions, stating that their effect "will be nothing less than to participate in the reconceptualization of the text, the library, and the act of reading itself" (Baker 1998: 283), and as Douglas Moffat pointed out in his introductory bibliographical review to a collection of essays (1998: 48), he was not alone:

McGann, Uitti and Greco, and Machan all seem to believe that computer technology provides, or will provide, the opportunity for their new visions of textual relations to be realized in something that will replace conventional codex editions. Traditional textual criticism developed along with and, so the argument goes, because of print technology and therefore reflects the nature of this technology; computer technology will enable a new, more accurate or comprehensive perception of the medieval manuscript culture to be realized.

(Moffat 1998: 48)

As Peter Robinson observed already in 1998, it had by the late 1990s become "a commonplace [...] for articles touching on any aspect of scholarly publishing and computing to observe that the impact of computing on scholarship in the late twentieth century is at least as great as was the impact of printing in the fifteenth century" (249). The prevailing belief seemed to be that "in the future most scholarly editions—documentary and critical—will appear in electronic form, probably accompanied by (or accompanying) printed texts published for general reading" (Williams and Abbott 1999: 85). In the mid-1990s, there also emerged concerted efforts to develop the methodology of digital editing, such as the Model Editions Partnership (MEP), which was established in 1994 to "explore what an electronic documentary edition should look like" and to "build some models to test various hypotheses about electronic editions" (Hockey 2004: 373).

After the initial period of excitement in the 90s, assessments of the possibilities created by the digital medium have over the last decade been markedly more subdued, tempered by "concrete experience of what happens to excited expectations when they come into hard contact with data structures, programming conventions, and a host of related choices we might not have even suspected we would have to make when we first embraced the idea of electronic or digital texts" (Searle 2004: 9). Consequently the digital format has not become the universal medium of scholarly editing, although the number of editorial projects with a significant digital component is increasing. It would seem that in the Anglo-American world, at least, the digital format has been embraced most comprehensively by documentary editors on the one hand, and Old and Middle English editors on the other. For

<sup>20</sup> At that time, the main candidate for realising this was the HyperCard software which had been included free with Apple Macintosh computers from 1987. HyperCard preceded the introduction of the World Wide Web and the HyperText Markup Language (HTML) markup language, developed by Tim Berners-Lee in late 1990. Although it was in many ways technically more advanced than the early versions of HTML, it was superseded by these new emerging technologies in the mid 1990s as they gained in both popularity and sophistication.

example Robinson (2006: 75) notes that all of the major editorial projects of Old and Middle English undertaken over the 1990s and 2000s “seem to have a significant computer component” (Robinson 2006: 75), while McGann (2004) points out that most digital editing projects seem to “have a strong documentary orientation” (382).<sup>21</sup> However, as Hockey (2004) noted ten years ago, the hype surrounding the topic of digital textuality and the speculation about the possibilities of the medium has tended to outweigh the actual practice of producing new kinds of digital editions.<sup>22</sup>

One of the major motivations for digital editing has been the perception that modern printed editions cannot accurately represent the features of medieval manuscripts, and have in fact “obscured manuscript records to an extraordinary degree” (Kiernan 2006: 263). While the primary aims of a digital scholarly edition are no different from those of a printed scholarly edition, “namely to make a given text available to an audience” and to articulate “the editor’s notions, perspectives, or theories” of what the text is (Vanhoutte 2006: 164), I will argue—following Pierazzo (2011: 475)—that contrary to what Tanselle (2006: 6) has claimed, digital editions are in fact ontologically different from printed editions and can use a much wider variety of means to accomplish these ends than printed editions.<sup>23</sup> This means that the principal challenge in the development of digital editing is to produce editions that are not restricted by the conventions of print editions and take advantage of the possibilities offered by the hypertextual digital medium to allow new kinds of scholarly interaction with the textual object (Fraistat and Jones 2006: 105; Kiernan 2006: 267). Editions that merely replicate the functionality of a printed edition in a digital environment are here not considered *digital* scholarly editions; if the use of the digital medium adds “no fundamental advantage [...] over the codex-based edition, it is better to stick to the book” (Vanhoutte 2006: 164).

While the transition from print to digital has often been compared to the shift from manuscript to print (cf. Robinson above), Chartier (1995) sees the former as “obviously more extensive than Gutenberg’s”, since it “modifies not only the technology for reproduction of the text, but even the materiality of the object that communicates the text to readers”, as well as “methods of organization, structure, consultation, even the appearance of the written word” (15), changing not only the way editions are *distributed*, but also the way editions are *made* (Robinson 1998: 249).<sup>24</sup> The situation is made simultaneously more complicated and more exciting

<sup>21</sup> Kline and Holbrook Perdue (2008) also urge “any documentary editor at least consider some form of electronic publication”, since many sponsoring agencies “insist on a strong electronic component in the projects they fund” (271).

<sup>22</sup> Partly for this reason, this thesis was from the beginning designed to prioritize the practical implementation of a digital edition as the means of exploring the possibilities of digital editing. The aim was to ‘put its money where its mouth is’, so to speak, basing the discussion of editorial principles and practices on an empirical basis in order to ensure it would not recommend practices that I was not willing to implement in practice.

<sup>23</sup> The ontology of digital editions will be examined in depth in chapter 5.

<sup>24</sup> As Bree and McLaverty (2009: 129) observe, also “the patterns of production and dissemination” of digital editions are likely to differ from those of traditional printed ones, in that they are “unlikely to be sold, like books, as individual items”, but will either be made freely available, or licensed to organizations as a part of larger electronic resources for a bulk fee. The significance of this view of the digital edition as a resource or a tool instead of a product is not limited to mere logistics, but also has a profound influence on the ontological identity of the edition itself, as will be shown in section 5.3.

by the fact that we are still in the middle of these changes; the possibilities of digital editions are “immense but frustrating” precisely because the practical conventions for producing them have not yet been established, “nor are such conventions likely to assume a stable form in the near future” (Crane 2006: 289):

Editors may justly feel that electronic editions have translated them from a stable environment with difficult but well-known problems into a river of Heraclitean flux, in which everything is changing from moment to moment, and the editor and edition are expected to adapt actively to those changes from moment to moment, without being able to rely on many of the principles which used to be stable guides to editorial thinking. Successful engagement with the mutability of electronic editions and the flux of the environment—surviving the experience of being tossed unceremoniously into the river—may require a different way of thinking about editions and the choices they embody. (Sperberg-McQueen 2009: 30)

### 4.2.1 Benefits of digital editions

On the conceptual level McGann (2004) has argued that representing originally paper-based textual objects in a digital form “entirely alters one’s view of the original materials”, because while “using books to study books constrains the analysis to the same conceptual level as the material to be studied”, digital tools “raise the level of critical abstraction in the same way that a mathematical approach to the study of natural phenomena shifts the theoretical view to a higher (or at any rate to a different) level” (383).

On a more practical level, various scholars have identified a number of specific ways in which the digital medium allows us to overcome the limitations of the traditional print edition. Already in 1995, Pichler (1995a: 770) summarized the advantages of digital (or “machine-readable” as he calls them) texts in general as: 1) “[...] easy, cheap, space saving and fast production, reproduction and distribution”; 2) openness to “revision, including corrections and additions, change of format, font and style etc.”; 3) amenability to “all types of computer assisted analyses, be they statistical, grammatical, stylistic etc., or content analyses”; and 4) the ability to “be converted into paper printouts and book editions” much more easily than paper editions can be converted into a digital one.

More than a decade later, Kytö, Walker and Grund (2007: 70), writing from a more corpus-linguistic viewpoint, very similarly list the main benefits of the digital format as the ability to 1) “include more material than is normally found in editions”, 2) provide information about and reproduce some characteristics of the manuscripts, and 3) perform computer searches on the material, making it easier for researchers to exploit the material. From the point of view of documentary editing, Kline and Holbrook Perdue (2008) see the digital edition as offering the user “1) access to several versions or views of the same text (image, diplomatic transcription, clear text), 2) options for organising and reorganising the edition’s texts in more than one way, 3) annotation that serves more than one document without cross-references or back-of-book index citations, 4) easy and immediate access to a cumulative index for an ongoing series, and 5) inexpensive and easy in-

corporation of maps, drawings, and other related images” (273).<sup>25</sup> For the purposes of a more detailed discussion, the benefits of digital editions mentioned above have here been grouped under five separate but interrelated topics, discussed individually below:

- 1) unlimited potential for inclusion of material,
- 2) ability to represent variance and multiplicity,
- 3) facilitation of analysis and contextualization,
- 4) flexibility and manipulability,
- 5) extensibility and fostering of collaboration.

### Unlimited inclusion of material

The first benefit of the digital medium is in many ways the most trivial. Because of the static two-dimensional nature of the printed page, a printed edition is always a compromise between the amount of information provided about the original textual object and legibility, and many of its editorial conventions are designed to maximize the presentation of information without making the pages of the edition “an incomprehensible mass” (Kline and Holbrook Perdue 2008: 144). The limited capacity of the printed edition manifests itself in two dimensions. In addition to the limitation in the ‘depth’ of information about the textual object that can be included on the printed page, i.e. documentary features like the material paratext as well as analytical and interpretive information, the amount of material is also limited in its ‘breadth’, e.g. in the number of manuscript versions of a work that can be represented. As Edwards and Moffat (1998) observe, for example the paucity of information provided by critical editions on variant manuscript witnesses is “an inevitable consequence of print technology and the economics of publishing” (Edwards and Moffat (1998: 229).

These limitations are partially the result of the physical format, and partially of the economics of print production. Already in 1995, Lavagnino saw computer storage as the solution to the problem of the printed editions’ cost increasing drastically with their size (112), and as recently as 2011 Pierazzo considered the reduced cost of publication as one of the major reasons for the current popularity of digital editions, alongside with “the greater possibilities of representation which the digital medium allows” (464). As a number of scholars (e.g. Johansson 2004: 95, Deegan 2006: 361, and Bree and McLaverty 2009: 127) have pointed out, digital editions are not constrained by the physical dimensions of the printed page or the economics of print publication, but rather by the imagination of the editor, and of course, the time available for the preparation of the edition.

<sup>25</sup> Although many digital editions—such as the Chaucer editions by the Canterbury Tales Project (Robinson 1996; Solopova 2000; Robinson 2004; Thomas and Bordalejo 2006) and many linguistic corpora—have been published on CD-ROM, Grenier-Winther (2004: 208) has noted that despite being digital, the disc is still a *closed* publication format, a product much like a printed book, and moving to online publication over the World Wide Web would provide even more advantages, namely: 1) the relatively unlimited scale of the edition, referring both to the amount of material and the number of alternative views on it; 2) the organic and dynamic nature of the edition, referring both to the addition of new material and to its display to the user; 3) the relatively universal and immediate access to all iterations of the edition, referring both to user access and independence from a publication and printing cycle; and 4) the possibility for interactivity with the reader, referring both to immediate query-result operations and to the possibility of user feedback and contribution.

This also has implications for the editor's role, which were being discussed already in the early 1990s. For example Brockbank (1991) noted that the "editorial service" demanded by digital editions "differs in kind from that traditionally found appropriate in the book" (103). While the constraints of the printed page force the editor to "select, sample, exclude, and to decide (often silently) between alternatives", the digital editor "is not required by the format to impose boundaries upon his choice of materials", but the overwhelming opportunities afforded by the medium place greater demands on "his capacity to offer creative guidance" to the reader by providing her with "apt data" while "keeping [the screen] clear of irrelevancies" (103).

This description highlights one of the central traits of the digital edition that will frequently emerge in this discussion, namely its tendency to *postpone* the decisions that have traditionally been seen as the core of editorial work, often to the point of relegating them to the user (Grenier-Winther 2004: 190). Instead of filtering the information provided to the reader about the textual object at the time of editing, the digital medium allows the editor to include a theoretically unlimited amount of information and to let the reader decide which of this information is relevant for her purposes. The editor's critical task, separating her from a mere hoarder of data, then becomes one of organising and labelling the data in a way that allows the user to make informed decisions about its relevance, i.e. presenting it in a structured manner and providing it with the necessary metadata.<sup>26</sup>

### Representation of variance and multiplicity

The ease with which multiple versions of a text could be published electronically on the Web accelerated the movement among textual editors away from the creation of a single, 'ideal' text - the only practical achievement for most scholars whose results would appear on a printed page. (Kline and Holbrook Perdue 2008: 27)

The advantages of digital hypertext—consisting of textual elements connected together by explicit and computationally processable links that encode various kinds of relationships between them<sup>27</sup>—are not limited to an increase in the *quantity* of text that can be included in an edition, but also provide new ways of expressing relationships between textual elements. The traditional textual topology is one-dimensional: textual elements are related to each other in terms of their distance and order, making text linear and directional. The concept of *linking* introduced by hypertext allows us to define additional lateral relationships between textual elements that are independent of their distance or order in the traditional linear topology, essentially creating a two-dimensional or net-like ('rhizomatic') textual structure:

<sup>26</sup> Even for an editor subscribing to the 'strong' editorial position of traditional critical editing, there is no longer a need to choose a *single* critical argument for exclusive inclusion; a digital edition can not only present several critical interpretations of the text, but can—and should—also present the reader with the evidence—in the form of documentary transcriptions of all manuscripts—that form the basis of the critical interpretations and allows them to be evaluated by the reader (Bodard and Garcés 2009: 91).

<sup>27</sup> For a more detailed theoretical discussion of the concept of *hypertext*, see e.g. the classic treatments by Bolter (1991) and Landow (1992, 1997).

The two salient features of all hypertext implementations are the powers to *gather* a system of texts into a stack and to integrate that system by making *links* among those texts. These links challenge the hierarchy of page order or volume organization by enabling rapid movement, either in vertical jumps forward and backward in the same text or in horizontal jumps from text to text. (Sanders 1995: 127)

Furthermore, since these links can be provided with explicit metadata identifying their structural or semantic significance,<sup>28</sup> they allow for the unlimited definition and representation of new relational dimensions.<sup>29</sup> This phenomenon of multidimensionality can also be viewed on the level of the physical page. Whereas the printed page consists of a single two-dimensional plane and consecutive pages are conventionally considered to form a linear sequence,<sup>30</sup> hypertext frees us from the physical constraints—or the “hard structures” (Bolter 1991: 41)—of the printed book and allows us to layer all the variant versions of the work onto each other using linking, making it possible to view in the text in two dimensions, either lengthwise, following the consecutive elements (words, paragraphs, pages, etc) of a single version, or ‘depthwise’.

As textual scholars and editors became aware of the properties of hypertext in the early 1990s, they saw it as a means of freeing themselves from the necessity of constructing a single text based on a textual tradition of variant versions. For example Lavagnino (1995) envisioned hypertext editions that do not assume a single established text but are “prepared to deal with more than one version of a work—with variants of individual words and passages, or with whole variant texts” (110). Already in 1999, Williams and Abbott observed that many electronic edition projects were making use of the possibilities offered by the digital medium and hypertext by combining diplomatic transcripts of the individual witness texts with a critical edition based on them.<sup>31</sup> As Bree and McLaverty (2009) point out, this means that the question of selecting a ‘copy-text’ no longer matters, since sev-

<sup>28</sup> This is in contrast to the implicit nature of traditional textual organization; we are used to assuming certain kinds of relationships based on the distance and order of textual elements (codified and studied as syntax and discourse structure), and it is difficult to redefine these implicit relationships without compromising communicative efficiency.

<sup>29</sup> As commonplace examples of these kinds of relationships we can mention for example the index of a printed book, which links together—however clumsily—points in the text that deal with the same concept, the collation of manuscript versions, where we essentially link words that we consider to be ‘parallel’ or somehow different versions of ‘the same’ word, and part-of-speech annotation, which essentially links together all words that we consider serving a similar syntactic function, regardless of their situation in the text.

<sup>30</sup> Although the pages of a codex are physically organized as a ‘stack’, the third line of a page is not usually considered to be in any special relationship to the third line of the next page, as the topological relationship between adjacent pages is seen to be one of latitude rather than altitude, the next page being conceptually located below the previous one, not underneath it. This traditional topology of the codex is strikingly illustrated by the famous early example of a paper *cybertext* on paper (Aarseth 1997) by Raymond Queneau, *Cent mille milliards de poèmes* (Queneau 1961) which consists of ten sonnets with identical rhyme schemes and rhyme sounds printed on ten consecutive pages cut into fourteen strips, with a single line on each strip. This places the ten lines in each of the fourteen ‘stacks’ into a paradigmatic relationship with each other, allowing for the generation of 10<sup>14</sup> different variations of the poem. However, it should be noted, that because of the physical limitations of the codex form, the use of the sequence of pages in an unorthodox way to represent variation or ‘depth’ instead of sequentiality or ‘length’ effectively limits its sequential length of the text to a single page.

<sup>31</sup> As an example they mention the *Piers Plowman* electronic archive, whose editors “plan for it



eral or all of the versions of the work can be included in the edition and linked to each other. This allows the user or reader—provided with suitable tools—to compare the texts using any one of the texts as a control for the others, automatically generate various kinds of collations and critical apparatus representing the whole range of variation, or generate quantitative measures of both local and global similarity and difference (Lavagnino 1995: 118-9; Bree and McLaverty 2009: 135).

While editions focusing on the representation of variation between manuscript versions, have—with the notable exception of the editions produced by the Canterbury Tales Project<sup>32</sup>—so far been produced mostly of printed modern literary works<sup>33</sup> several scholars<sup>34</sup> have noted that the ability of hypertext to deal with variation makes it is especially well-suited for the representation of medieval textuality, offering a “technology to construct editions that respond to some of the documentary and textual diversity of Middle English works” (Machan 1994: 190-1). As Greetham (1998) has put it:

Hypertext has made visible what toilers in the fields of medieval textuality have generally accepted but been unable to chart effectively within the constraints of the print medium: that intertextual penetration and consubstantiation is the norm, not the aberration, in medieval textuality, and that postmedieval (especially modernist) essentialist theories of genre and form cannot adequately represent the documentary pluralism, fragmentalism, and antidisciplinary cross-fertilizations of the postmodern and the premodern. [...] In a wonderful cultural and historical hermeneutic circle, it has taken the destabilizing electronic configurations of hypertext to embody the medieval condition of *mouvance*, and it is therefore no surprise that hypertext has been seized on by medievalists as a particularly fruitful medium for the productive display of the ‘web’ we have felt to be the presiding figure of medieval textuality. (Greetham 1998: 298-299)

The problem of representing parallel versions is especially acute in the case of works whose versions differ significantly in their organization, which cannot easily be presented in a parallel edition without prioritising the textual organization of one version and obscuring that of the others (Blake 2000: 33; Hockey 2004: 362). Hypertext has also been argued to be a much better representation of the highly

---

to include not only digitized color images of the fifty-four manuscripts (with codicological and linguistic description of each) and TEI-conformant transcriptions of these manuscripts, but also editorially constructed texts of the A, B and C scribal archetypes and critical texts of the three authorial versions” (Williams and Abbott 1999: 84). According to the web site of the project (<<http://www3.iath.virginia.edu/seenet/piers/>>), six manuscript versions have been published (on CD-ROM) so far, the latest in 2008.

<sup>32</sup> See e.g. Robinson 1996, Solopova 2000, Robinson 2004 and Thomas and Bordalejo 2006.

<sup>33</sup> Examples of these kinds of editions include the SGML-encoded edition of Coleridge and Wordsworth’s *Lyrical Ballads* by Bruce Graver and Ronald Tetreault (2011), the genetic edition of Stijn Streuvels’ novel *De teleurgang van den Waterhoek* by Marcel de Smedt and Edward Vanhoutte (De Smedt and Vanhoutte 2000), and the digital historical-critical edition of the collected writings of one of the earliest Finnish novelists, Zachris Topelius, by the Svenska Litteratursällskapet i Finland (SLS 2008-). All of these editions, while slightly different in their underlying technology, allow the reader to view all published (and some manuscript) versions of the work side by side, highlighting the readings in which they differ.

<sup>34</sup> See e.g. Cerquiglini 1989: 116, Bolter 1991: 86, Hanna 1992: 122, Machan 1994: 190-1, Sanders 1995: 129, and Greetham 1998: 298-9.

intertextual, compilatorial and commentatorial nature of medieval textuality than the “stand-alone or ‘unitary’ text” of printed editions, “designed to produce the illusion of a continuous, fully coherent and complete reading intended [...] to be performed unidirectionally (left-to-right, top-to-bottom, and beginning-to-end)” (Sanders 1995: 126). As Carlquist (2004: 106) has also argued, this is especially true of composite manuscripts and other manuscripts that are not intended for sequential reading—such as recipe collections—as they are very difficult to represent in paper editions in a way that would allow their receptional characteristics to be highlighted.<sup>35</sup> Digital editions, which allow linking and the fluid recombination of different manuscript items in varying sequences, make it easier to represent the ways in which these kinds of texts were intended to be read.

### Facilitation of analysis and contextualization

Digital scholarly editing offers the chance to organize paratexts and transmitted material in much more dynamic and complex manners than is possible within the printed edition. The modular, database logic along with the potential qualities of digital media [...] push the edition towards becoming an *archive*. (Dahlström 2009: 40)

In addition to links between related textual elements, digital texts can also be linked to a variety of other kinds of metadata about the text and its elements. While the addition of annotation is also possible in printed texts, the fact that digital texts are not read directly but mediated by a computer (Bolter 1991: 42-3) provides them with two significant and interrelated advantages: 1) metadata can be ‘hidden’ from the user so that it does not hinder the reading of the text itself, allowing for the addition of unlimited amounts of metadata, and 2) the computer can dynamically manipulate, analyse and display the text in different ways based on this metadata. The fact that both the text and the metadata describing it are encoded in a machine-readable way allows linking to be applied not only between textual elements but also between textual elements and metadata describing them. This offers us unprecedented potential for the computerized analysis of not only the complex interrelationships between textual elements but also those existing between textual elements and various kinds of descriptive and analytical metadata.

Already in 1989 Cerquiglini envisioned a digital edition that would include “great many minor bits of information, which should remain virtual so they will not get in the way of reading but which one needs to be able to locate: makers of concordances, frequency lists, tables of rhymes, every sort of calculation, codicological and paleographic data, and so on—everything that a printed edition usually abandons or from which it makes a painful choice” (Cerquiglini 1999: 80). Almost ten years later, after the production of the first generation of digital editions and soon after the introduction of the World Wide Web, Baker (1998) saw even wider applications for the hypertextuality of digital editions, namely “the ability to link

<sup>35</sup> Although he does not use the term, the kinds of texts Carlquist describes in his article—collections of sermons, hagiographies and law texts—are very similar to recipe collections in that they belong to the category of *discourse colonies* (Hoey 1986, 2001), described briefly in section 8.2.

the text into a web of information that includes related texts, commentaries, dictionaries and other reference works, manuscript facsimiles, artwork, music, and videos”, turning digital text into “one element in an electronically replicated cultural context that, in theory at least, may be as open-ended as the real-world context on which it is modeled” (264).<sup>36</sup>

While most scholars envisioning and developing digital editions over the last two decades have focused on the benefits of integrating analytic and contextual data about the text to digital transcriptions, less attention has been paid to the use of digital hypertext to include more information about the *document* itself. Furthermore, this discussion has mostly focused on the ease and benefits of including digital facsimile images in digital editions as a means of helping “modern readers negotiate a culturally strange text in its original form with all its foreign formats, including its codicology, paleography, scribal abbreviations, shifting orthography, unexpected word divisions, unknown word-hoard, missing metrical layout, and lack of any helpful punctuation or capitalization” (Kiernan 2006: 263). The inclusion of detailed digital images is definitely a major boon for digital editors and does allow the reader to evaluate the documentary nature of the edited text (Reimer 2004: 172) and to ascertain visual details like “where a note falls on the page, whether the hand changes from one note to another, and whether the editor has made any errors in the transcription” (Jackson 2004: 79).<sup>37</sup> However, for a corpus linguist they have a critical shortcoming, namely that they are not searchable in the manner of digitized text.

The property of digital editions that makes them more powerful as research resources than traditional printed editions is the fact that digital data can not only be easily stored and transmitted, but also *processed*—“mathematically sorted through to show hidden relationships, new arrangements, different views, and expanded, contracted or concatenated knowledge” (Terras 2010: 50-1). The minimum requirement for an edition to be used for modern corpus-linguistic research is naturally its machine-readability, the principal defining quality of a digital edition. However, as was pointed out in subsection 4.1.2, the variationist corpus-linguistic method is based on relating the occurrence of linguistic features to various kinds of contextual features, and this kind of analysis can benefit immensely from the presence of metadata describing the documentary features of the text. This kind of descriptive metadata can help broaden the scope of the linguistic assessment of the text by taking into account multiple communicative modes, and is especially useful for pragmatically oriented discourse studies.<sup>38</sup>

Although these possibilities have been recognized by many scholars, Buzzetti

<sup>36</sup> The explosion of the Internet into a ubiquitous phenomenon affecting all aspects of our culture over the last two decades has also opened up the possibilities of digital editions, as it allows editors and scholars “to bring together extensive corpuses of primary materials” (Palmer 2004: 351), situating a work “in a nexus of social, contextual, and historical materials, all of which contribute to the totality of its meaning” (Deegan 2006: 358).

<sup>37</sup> The unavailability of high-quality digital images for the manuscripts edited here, and the prohibitive copyright fees for the reproduction of the existing microfilm facsimiles have precluded the inclusion of digital manuscript images in the present edition even for this purpose.

<sup>38</sup> For example explicit structural metadata identifying the different parts of a text allows the corpus linguist to determine whether a certain linguistic feature occurs more frequently in a specific part of the text (e.g. introductions, beginning or end of chapters, headings) than in others. Similarly, metadata representing the visual paratext (e.g. typography or layout) can reveal what kinds of linguistic structures are being visually highlighted in a text.

(2009: 45) has argued that most present-day digital editions in fact “do not fully exploit the distinctive features of the digital form of textual representation to obtain better critical and analytical results” and to “improve our critical engagement with the text through effective computational processing”. According to Buzzetti, the main reason for this is the tendency of humanities scholars to see the addition of metadata to their texts in the form of markup and textual annotation as something belonging to the field of computer science and thus outside their own expected competence. However, the annotation of digital texts with metadata, whether describing the original document or analysing the linguistic or other features of the text, is in fact merely a form of reporting the results of philological and linguistic research in a way that makes it usable as the basis of further research, and should therefore be at the core of our scholarly practice (see section 5.4).

### Flexibility and customisability in use

The difference between a book edition and an electronic edition lies, however, in an electronic edition’s potential to be able to make the types of interpretation - and their differences - explicit and extractable, to give the users the possibility to choose between the different levels of interpretation, and to realize them in different ways: e.g.: to choose an ad literatim printout rather than a normalized one, but at the same time have text, which was originally underlined, printed in italics. With regard to these demands a machine-readable version has considerable advantages over a book edition. (Pichler 1995b: 695)

As Sperberg-McQueen (2009: 30) points out, a large part of editorial theory throughout its history has been concerned with identifying and weighing the relative merits of various choices among alternative editorial approaches that have been considered mutually exclusive. While a printed edition requires editors to choose at an early stage one particular view into the original manuscript text, deciding what information to present to readers and how to present it, a digital edition allows some of the choices on the former and most of the choices on the latter to be made not by the editor, but by the user according to her needs—although obviously curtailed by earlier decisions made by the editor (28–30). This idea of presenting texts dynamically according to the user’s needs was envisioned already in 1989 by Cerquiglini, who saw the digital edition as accommodating “varied textual masses, which the reader consults by making them appear in different ways on a computer screen” (79), and has frequently been seen as one of the major benefits of digital editions over paper ones<sup>39</sup>

One reason why this feature has been seen as so important is given by Lavagnino (2009: 65), who observes that editors of English-language texts are very often preparing editions simultaneously for (at least) two very different audiences. On the one hand they are editing for other editors and textual scholars, towards whom much of the apparatus surrounding the text is geared, and on the other to “common readers”. These two audiences have very different needs, and since there is no single ‘correct’ way to edit texts—the ‘correct’ approach always depends on

<sup>39</sup> See e.g. Baker 1998: 272, Williams and Abbott 1999: 101, Driscoll 2006: 258, Van Hulle 2006: 258 and Parker 2006: 207.

the purpose of the edition—a single all-purpose representation of a text is an impossibility (Shillingsburg 1991: 26-7). The most obvious way of dealing with this problem of multiple audiences is, of course, “to provide a different edition to each audience, varying the nature of the textual discussions and apparatus” (Lavagnino 2009: 70), providing the textual scholars with a critical edition, a diplomatic text, a facsimile reprint or simply a “demonstration of some textual theory”, and the general readers with “a reasonably honest and readable text with appropriate apparatus” Pearsall (1994: 117).<sup>40</sup> However, the digital edition offers a more economic and elegant solution to this problem:

Because delivery software can filter the body of information contained in the edition at the time of reading, the same edition can be used to generate the forms of presentation thought most useful to beginning students, or to specialists with various interests. [...] A multi-purpose edition can be created simply by including all the information needed for each purpose (typically there will be more or less overlap) and providing a suitable user interface for each.

(Sperberg-McQueen 2009: 33)

This means, as for example Pichler (1995a: 770) and Dahlström (2009: 28) have observed, that the digital edition and the traditional printed edition are not strictly alternative to each other, since the ability of a digital edition or archive to output various document forms of its textual material means that it can also be used to produce various kinds of printed editions embodying different theoretical viewpoints and suited for different purposes. This generative relationship makes the traditional paper-based edition essentially a derivative sub-set of the multi-dimensional digital edition (Grenier-Winther 2004: 210-1). This also raises an interesting issue with regard to the traditional editorial typologies (see subsections 3.1.1 and 3.2.1). As Pierazzo (2011, 2009) points out, while editions like *Jane Austen’s Fiction Manuscripts Digital Edition* (Sutherland 2003) or the present edition would traditionally be characterized as ‘diplomatic editions’, this is in fact slightly misleading since the diplomatic rendering of the manuscripts is only one possible output among others that are “contained simultaneously *in potentia* within the same source file” (474):

Textual editions based on digital encoding can, for instance be easily presented on a website with different layout formats (readable version, diplomatic version, semi diplomatic version, and so on), some of them even offering to users the possibility of building their own visualization of the text.

(Pierazzo 2009: 170)

This “proteiform, cumulative nature of digital editions” (474) thus challenges both the whole concept of ‘an edition’ as a product (see chapter 5), and the commonly accepted categorization of editions based on the level of editorial intervention and treatment of textual features.

<sup>40</sup> Hanna (1992: 122) has seen this same conflict of interests in a slightly different light, as a conflict between the requirement of modern readers of a singular, unified text, and the multiple reality of the manuscript text, but proposes a similar solution, namely the production of “a range of use- or interest-driven possible editions” that would not only serve different interests, but also “approximate, through diversity of approach and method, that plurality which is a property of its subject, texts in manuscript” (Hanna 1992: 122).

### Extensibility and fostering of collaboration

A burst of collaborative effort has been a surprising bonus of the humanists' entrée into the electronic domain. (Eaves 2006: 219)

Unlike printed editions which—once printed—are “fixed and lasting”, a digital edition published on the Internet “is flexible and can easily be adjusted” (Bøe, Jørgensen and Taugbøl 2004: 60). In addition to the usual provision of errata, online digital editions can also be extended in several different dimensions (Lavagnino 1995: 115). First of all, new versions of the work—such as newly discovered manuscripts—can be added to the edition, secondly, the edition can be linked to compatible editions of other works, forming digital libraries or corpora, and thirdly, new metadata can be added to the existing texts, enriching the analytical capabilities of the edition. In cases where the content of the edition is separated from the tools used to present and analyse it, the capabilities of the edition can also be extended through new tools developed after the preparation of the edition, allowing the development of analytical tools to follow the development of the (initial) edition itself, and thus be based on the actual needs of scholarship (Lavagnino 1995: 115).

The ability to extend a digital edition is not limited to the original editors, but the digital environment also allows the creation of editions that they can be developed further by other scholars or used as raw material for new kinds of resources.<sup>41</sup> While the idea of collaborative work and building on the work of previous scholars is a central principle of scientific research, the digital environment, by enabling scholars to ‘publish’ and share the results of their labours themselves without the financial and infrastructural overhead of a traditional publisher (Robinson 1998: 253) has enabled the sharing of research materials and other resources with unprecedented ease. In his history of corpus linguistics, Johansson (2008) argues that the success of corpus-based research methods in English linguistics is largely the result of many early corpora being placed freely at the disposal of the international scholarly community.<sup>42</sup>

In order to maximize the usefulness of the time and scholarly effort spent on producing a digital edition, they should be designed and published in a way that does not merely provide a finished product to a reader or a user, “but may also benefit the next generation of editors or current editors with different goals” (Ore 2009: 115). One of the basic ideas behind this thesis and the accompanying edition is that digital editions that are prepared using established and open technical standards, clearly documented and published under a suitable Open Source license

<sup>41</sup> This idea of digital editions as dynamic ‘work sites’ rather than finished ‘products’ is discussed in more detail in chapter 5.

<sup>42</sup> According to (Johansson 2008), this sharing of resources was “an important and rather novel aspect of the development of corpus linguistics”, (35) and for example “the world-wide importance of the Brown Corpus stems from the generosity and foresight shown by the compilers in making the corpus available to researchers all over the world” (38). In the field of historical corpus linguistics, the *Helsinki Corpus of Early English Texts* (HC 1991), which has always been freely available through the University of Oxford Text Archive (<<http://www.ota.ox.ac.uk/>>) has gained a very similar status, producing a large body of research and even several parsed corpora based on its component parts (Kroch and Taylor 2000; Taylor et al. 2003; Kroch, Santorini and Delfs 2004).

could allow editorial work to become cumulative, allowing them to be “improved, enhanced, and made more accurate by future researchers, and the apparatus based on them [...] modified rather than rebuilt from scratch” (Parker 2006: 204):

The idea that once an edition was done it was unlikely to be done again is not a product of the digital age. Once text is digitized, particularly when using markup like XML, it becomes far easier to re-purpose it, run it through text analysis tools, add new levels of encoding, and open up the possibility that other scholars might find new uses for our old editions. (Hajo 2010)

### 4.2.2 Criticism of digital editions

Despite the numerous benefits outlined above, digital editions have also been criticized on various grounds. In the mid-1990s, the most common objection to digital hypertext editions was that they were difficult to read in comparison to printed editions. However, as Lavagnino (1995) pointed out already at the time, this was actually not an objection to the editions themselves, but rather to the computer hardware of the era, “to the fact that our screens are too small, and less pleasant to read from than books are” (113). While technically justified at the time, this objection was quite correctly understood by many scholars (e.g. Bolter 1991) to be only a temporary one, being addressed by developing technology.<sup>43</sup> In addition to being ephemeral, this criticism can also be argued to be beside the point; most digital scholarly editions are not really intended to offer texts for casual reading but rather “functional objects to be used” (McLoughlin 2008: 4) as an “aid to scholarly labor” (Lavagnino 1995: 113). A more recent variation of this argument that has occasionally been presented against the addition of large amounts of metadata to digital editions or corpora is that adding any kind of *markup*, whether implemented in XML or ad-hoc codes, makes the text “basically ‘unreadable’ for the human eye” (Claridge 2008: 253). This argument, however, ignores the *mediated* nature of digital text, mentioned above. While for example raw XML code can seem ‘unreadable’ to someone not used to it, it is not intended to be read ‘as such’ but rather visually formatted using a suitable style sheet.<sup>44</sup>

#### Lack of tools

One significant hindrance to the proliferation of digital editions with rich metadata—and especially of research making use of such metadata—has been the lack of user-friendly and accessible tools for the manipulation and analysis of digital editions. Already in 1995: 113 Lavagnino argued that while the inclusion of images and transcriptions of every version of a work and the ability to display any two versions side by side—which was part of the core functionality of every hypertext edition in progress at the time—“was indeed essential, [...] it was also not enough”

<sup>43</sup> This objection can be seen to have lost its validity at the end of the last decade with the introduction of eBook readers like the Amazon Kindle (introduced in 2007) and tablet computers (the first generation of Apple’s iPad was introduced in 2010). At present it is actually *easier* to read a digital text virtually anywhere than it is a printed book beyond the size of a small booklet.

<sup>44</sup> In some ways, calling an annotated text ‘unreadable’ is the same as looking at a plain text in an editor which shows it as a series of hexadecimal values and calling it ‘unreadable’.

(115). What is needed in addition to these new kinds of editions are “analytical tools that scholarship will demand for working with multiple versions” (115). The question of tools that allow the efficient display, searching and analysis of the wealth of information provided by metadata-rich digital editions seems to be a ubiquitous one. Even before digital editions were being prepared, Cerquiglini (1989) saw the ability of a digital edition to visualize and analyse the connections and dynamics of the variants discovered and marked by the editor as even more important than the computer’s ability to display several variants of a text simultaneously. In the mid-1990s when the first generation of digital editions were being published, Eggert (1994) was still waiting for the time when “satisfactory software for electronic editions is finally written” (7), and only a few years ago (Hajo 2010) called for the digital editing community to develop “better tools and encoding environments to win over editors and other content providers” to XML-based digital editions and argued that “we need to make available the programs and stylesheets that will make these texts display clearly and will take advantage of the encoding to generate valuable searches”, implying that there still is at least a perceived shortage of such tools.<sup>45</sup>

As Lavagnino (1995) has observed, analytical tools for working with metadata-rich multi-version texts are not only more difficult to develop than basic systems for hypertextual display,<sup>46</sup> they are also “less likely to be provided by commercial vendors, who are not going to recognize on their own why they are necessary or perceive any large markets for them” (116). This leaves the task of tool development largely to dedicated scholars, which in turn tends to result in tools tailored to the needs of specific projects. This means that digital editions are currently published mostly within a dedicated research environment or user interface, either online or on CD-ROM, and cannot easily be used outside of it, which has the unfortunate side effect of limiting the use of that edition to the purposes envisioned and enabled by the developers of its interface.<sup>47</sup> Thus the perceived lack of tools for using digital editions is simultaneously real and illusory. While it is true that there are practically no ready-made general-purpose tools for the viewing, searching, and analysis of digital editions, there do exist a variety of high-level application environments that allow for the flexible development of even extremely sophisticated research environments with very little knowledge of the underlying programming languages or other technical aspects of software development. Again, the problem is that the development of such environments usually takes place in the context of a specific editorial project and is oriented towards the specific needs of that project.

<sup>45</sup> Fortunately, as mentioned in subsection 4.3.4, the *Textométrie* project at ENS Lyon is fortunately making good progress in remedying this source of criticism by developing a general-use linguistic analysis software for richly annotated digital editions and linguistic corpora.

<sup>46</sup> Lavagnino (1995: 116) lists four features that are required of these tools: 1) *selecting* versions to look at; 2) *comparing* versions; 3) *constructing* new and possibly more representative versions of the text on the basis of the information available; and 4) *integrating* all this study with other scholarship and criticism. From a corpus-linguistic point of view, the list could be supplemented with the ability to make searches on the text based on both its textual content and the included metadata.

<sup>47</sup> Furthermore, the lack of resources often leads digital humanities projects to “adopt and adapt tools that were originally developed for other needs and audiences”, which means that it can be difficult to create a ‘slick’, unified and user-friendly interface for the research environment (Kirschenbaum 2004: 539).



The solution advocated by this thesis, described in section 5.1 and demonstrated by the present edition, is based on separating the creation of data resources (i.e. digital editions) from the development of tools for their analysis, and focusing on developing a standardized *interface* or a strictly specified data structure through which any tool implementing that interface can access any edition that provides its data in the format defined by the interface. This would allow those digital humanities scholars who have an inclination towards software development to focus on the process-oriented development of tools that would be usable with a large number of editions developed independently of each other but implementing the same interface.<sup>48</sup> By defining and documenting a strictly restricted set of the encoding and annotation conventions, based on the relatively loose standard defined by the Text Encoding Initiative, this thesis seeks to take the first steps towards such an interface for linguistically oriented editions of historical texts.

### Loss of editorial control

While the flexibility and customisability that digital editions afford their users are usually considered to be a positive thing, some scholars have also pointed out that the fluidity that it implies for the edited text “can cause serious problems at the same time as it conveys many benefits” (Deegan 2006: 361). While the ability of the user to reorganize the material contained in the edition makes the organization of the edited material less critical for the editor (Kline and Holbrook Perdue 2008: 223), it can also be seen to lessen the control the editor has on his edition. For example Cooper (1998), discussing the Canterbury Tales Project, takes a rather cautious stance towards the freedom allowed by digital editions. Although she does concede that “it may yet be that electronic methods will recover ‘what Chaucer actually wrote’ where traditional approaches have failed”, but she worries that “by making the full range of manuscript readings and misreadings accessible to every Chaucerian critic, they might also allow a process of pick-your-own-text without any of the established editorial controls, such as could invite critical anarchy” (Cooper 1998: 83). Whether this ‘critical anarchy’ is considered a positive or negative state of affairs is very much a question of editorial philosophy and textual world-view: from the point of view of New Philology and historical linguistics, this lack of “editorial controls” can in fact be seen as an accurate representation of medieval textuality.

### Lack of status and stability

Although the fact that digital technology makes it easy for scholars to publish their editions over the Internet without the intervening infrastructure of publishers is generally considered a benefit (Robinson 1998: 253; Kline and Holbrook Perdue 2008: 28; Pierazzo 2011: 464), it has also been seen to involve its own problems. For example Roueché (2009: 168) raises the issue of “authenticity”, or the problem of establishing and authenticating the status of electronic online publications. Although their numbers have greatly increased over the last decades, the status of

<sup>48</sup> From a corpus linguistic point of view this would also have the added benefit of creating a pool of standardized digital texts which could be used for compiling corpora simply by organising them into principled subsets.

online publications like digital editions is still problematic in two respects. First of all, online publishing—if acknowledged at all in academic evaluation—is still considered to be inferior to traditional print publication, as for example Jensen (2004: 553) and Roueché (2009: 164) have noted.<sup>49</sup> Secondly, there is the problem of how to ensure the reliability of an online resource through some kind of a peer-review process and to communicate this information to its users. Unfortunately, both of these issues are mostly features of public perception, and as such beyond the influence of individual editors. If we wish to establish digital editions and other online research resources as valid outputs of academic work, the best we can do is to aim at producing resources that fulfil the strictest standards of academic rigour, and to work with reliable academic institutions in developing methods of certifying them as such.

Another problem of digital publication, pointed out by Eggert (2009), is one of longevity. Although he predicts that e-publication will become the primary format of publishing scholarly editions within a few years, “no general editor can ask scholars to spend several years of their life working on an edition if the stability of the reading texts they establish cannot be guaranteed beyond the lifetime of the software company and of the public funding (almost certainly temporary) of an electronic repository” (74). Discussing the status of the *William Blake Archive*, (Eaves 2006) acknowledges that it is available online for free only by virtue of the sponsorship of several institutions and recognizes the precarious nature of its existence and the uncertainty often faced by digital editors:

It would be foolish to assume that any of these resources is permanent, and we will never be a set of volumes sitting securely on library shelves. We plow forward with no answer to the haunting question of where and how a project like this one will live out its useful life.  
(Eaves 2006: 218)

It should, however, be noted that this institutional instability only pertains to the online *availability* of a digital edition, not to its existence as a set of data files and program code that can be preserved on a variety of storage media and archived in a library just like a printed book.<sup>50</sup> As Kirschenbaum (2002) has pointed out, the “opposition between fixed, reliable printed texts on the one hand, and fluid and dynamic electronic texts on the other—an opposition encouraged by the putative immateriality of digital data storage—is patently false” (24); both are subject to corruption, but neither undergo transformations on their own accord: the much vaunted ‘play’ of digital texts is “only screen deep” (26) and takes place on the level of presentation. This, as will be argued in chapter 5 and section 5.4, should be seen as a strong argument for the separation of the data from its presentation and for the use of recognized and well-documented encoding standards.

<sup>49</sup> This applies especially to interactive online editions which are rarely published by an established publisher but more likely by the sponsoring institution of the editor or even the editor herself.

<sup>50</sup> Whether this collection of files outside the active computing environment can be equated with ‘the edition’ is a matter of ontology and will be discussed in chapter 5.

### Increased editorial workload

While digital editions have above been characterized as being easier and more economical to publish and distribute than traditional printed editions, this characterization ignores the “considerable amount of work in tagging and encoding” that their preparation requires (Reimer 2004: 176). As Greetham (2004: 37) points out, the “romanticism of liberation technology has often conveniently repressed the confining, detailed, anally retentive *input* of electronic text, preferring instead to celebrate its superficial, phenomenological, fluid *output*” (37). However, whether they require more work or “are generally more expensive to produce than print editions” because of the expenditure involved “in technology and associated expertise”, as Bree and McLaverty (2009: 129) claim, is debatable.<sup>51</sup>

What is decidedly true, however, is the observation of Bree and McLaverty (2009) that digital editing requires new kinds of skills from the editor. Since many of the matters which in print editing were taken care of the publisher need to be accounted for by the editor herself, “[t]echnical expertise becomes as important as academic endeavour” (127). This is often a problem for humanities scholars, since many of them “still expect to define a problem, hand it over to a group of ‘technicians’ and receive a shiny new machine in return” (Roueché 2009: 165). As Roueché points out, this approach “has made many software manufacturers and consultants very rich, but it is not financially viable for non-profit-making enterprises” (165), and has the additional problem of making continued maintenance and development of the resource very difficult. The solution proposed by Roueché is to have editors work together with humanities computing experts to develop the technical solutions required by the digital edition, although Hajo (2010) argues that editors still need to learn at least the rudiments of the technology their editions make use of, because otherwise “we won’t be able to participate fully in decisions made on how the texts should be tagged, nor will we be able to fully explore the possibilities of digital editing”.

Finally, Sutherland (2009) has expressed scepticism about the real, practical usefulness and worth of electronic editions, pointing out that they are often being developed very much from the point of view of their producers, without sufficient thought into their eventual use. According to her, “we have not yet thought hard enough about who will use electronic editions or how often or for what real purposes” (19). While this may or may not be true for literary digital editions aimed at literary scholars and digital archives of historical documents aimed at historians, it seems clear that based on both its successes in shedding light on the history of English and the limitations of current historical corpora, the discipline of historical corpus linguistics is the perfect field of application for richly-annotated digital editions of historical documents.

---

<sup>51</sup> In terms of equipment, electronic editing requires no more than any modern scholar is using in any case, and the time and effort of acquiring the expertise required only has to be spent once, and is usually quite modest in comparison to the time and effort spent on the scholarly work itself. As an added bonus, it results in editors and humanities scholars with skills that are extremely portable for a wide variety of scholarly and practical tasks.

### 4.3 Editing for historical corpus linguistics

In a famous article (Tanselle 1978) listed the essential features that need to be retained in a diplomatic transcription of modern holographic documents (pp. 50-51), but such lists are neither possible nor desirable for digital editions: in fact the editor's goal needs no longer be 'to reproduce in print as many of the characteristics of the document as he can' (p. 50) but rather to achieve the scholarly purpose of the edition—a purpose which, by definition, varies.

(Pierazzo 2011: 475)

As the above quotation points out, the fact that the representative potential of digital editions are no longer limited by their medium means that their contents and features should be determined by their scholarly purpose. Even though the increased capacity and flexibility of digital editions, discussed above, means that they can accommodate a variety of uses, even digital editions cannot be all things to all men and commonly have a principal function that guides their design. As has already been established, the principal purpose of the present edition and the guidelines documented in chapters 10 and 11 is to represent the linguistic, textual and paratextual features of the original documents in a form which can be analysed using corpus linguistic methodology.

This section will examine the relationship of corpus compilation—focusing on historical corpora—to editing. This relationship is a rather complicated one: while practically no editions have been prepared with corpus linguistic research in mind until very recently, most historical corpora have been compiled from printed editions of historical texts. Considering that corpus linguistics requires computer-readable texts, this necessarily involves an additional process of digitization which further distances the corpus from the original documents. Digital editions, on the other hand, are very similar to textual corpora in practical terms, being different mainly on the conceptual level. Despite this similarity, which makes it quite easy to accommodate the requirements of corpus linguistics in the preparation of a digital edition, very few digital editions have been prepared with corpus linguistic study in mind, possibly because of being prepared mainly within either a literary critical or historical tradition.

#### 4.3.1 Compiling corpora from printed editions

While some historical linguists have “acknowledged and dealt with the problems of critical editions that narrow down a greatly varied manuscript tradition to one canonical text” (Grund 2006: 110), printed critical editions have nevertheless been used as the basis of historical corpora of English until very recently, their problematic nature having been largely obscured by the process of detaching them from their original printed form and including them in a corpus. As Dollinger (2004) has pointed out, the fact that for example the *ARCHER* corpus makes no overt mention of whether manuscripts or editions were used to supply the texts indicates that “the practice of using editions has somehow become paradigmatic in certain

areas of historical corpus linguistics” (6). Even quite recently, Johansson (2004: 95) pointed out that the idea of archives containing *manuscript* texts in digital form was still a novelty to many philologists and corpus linguists. However, in recent years historical linguists have, at least to some extent, recognized the importance of working with individual manuscript texts instead of critical editions (see e.g. Grund 2006: 110 and Claridge 2008: 250), and it seems likely that the use of edited texts as sources for corpora is largely based on their convenience to the compiler.

### Reasons for using editions

While aware of the superiority of original documents as linguistic witnesses, Claridge (2008) reminds us of the fact that the decision between using original documents or printed editions for compiling the corpus is an unequal one: in using original documents the compiler has to face the “challenges presented by the form of early texts”, while the use of editions relegates them to the editor (250). Hunston (2008: 157) has an even less optimistic view of the resources of corpus projects, pointing out that even texts that are available only in printed paper editions might be avoided because their scanning or keying might be beyond their resources. According to her, keying texts in by hand “is obviously very time-consuming and is generally avoided unless the texts concerned are unavailable in any other way” (158), which is why corpus compilers have been perhaps too quick to settle for whatever version of a text is conveniently available.<sup>52</sup> Even if the resources for keying in text are available, the use of original documents as the basis of corpus compilation is complicated and even restricted by the fact that manuscripts and early printed books are stored in various, often distant repositories, requiring travel and fieldwork in less than optimal conditions (Claridge 2008: 250). Furthermore, “linguists are not necessarily paleographers” (251) who would have the skills required for working with original manuscript documents. For anyone who has participated in the compilation of a textual corpus, it is therefore obvious that “[a]ll corpora are a compromise between what is desirable, that is, what the corpus designer has planned, and what is possible” (Hunston 2008: 156).

Although the exigencies of economy have played a significant role in the use of printed editions for the compilation of historical corpora, some scholars have also seen it to have inherent benefits. For example Markus (1997) obviously does not see the use of printed critical editions as detrimental to historical corpus linguistics, remarking that a “rigorous *ad-fontes* policy” would also ignore “what editors have achieved and are good for” (222). Claridge (2008: 250) summarizes the “obvious” advantages of the use of printed editions for compiling historical corpora as follows:

- 1) the texts are fairly easily available and do not require travel to remote libraries,
- 2) the editorial work with its numerous decisions has already been done,
- 3) the texts exist in a legible format which may even make scanning possible and in any case makes manual keying in easy.

<sup>52</sup> Speaking of modern corpora, Hunston (2008) is quite frank about the fact that “[w]ith the larger corpora that are expected today, [...] obtaining text in electronic form, either from a publisher or from the internet, is the optimum way of building a corpus of written texts” (158).

However, she admits that this approach has three disadvantages:

- 1) many editions have not been prepared by or for linguists, but rather by and for historians and literary scholars, whose concerns are not necessarily the same as those of linguists;
- 2) the linguistic decisions—such as spelling normalization—may not be adequately documented in the edition and are thus not documentable at the time of inclusion into the corpus although they significantly affect the results of linguistic research performed using the corpus;
- 3) editions may be copyrighted and thus restrict the availability of the resultant corpus.

Both of these lists highlight the significant role of practical considerations in the compilation of corpora and the fact that printed editions are used mainly as a means of facilitating the compilation process and overcoming some practical problems, even if it also introduces some new complications. However, as Hunston (2008: 157) pointed out, the digitization of printed editions for a corpus involves a significant amount of work and in a sense duplicates a lot of the work done by the editor compared to the situation where the edition is already presented in a digital form. Thus, from a purely practical point of view, an optimal solution for the compiler of a historical corpus would be to use well-documented and linguistically sound digital editions which would provide her with all the benefits of using printed editions without most of the drawbacks, and eliminate the need for work-intensive digitization.

### Problems of using editions

In addition to the practical issues of corpus compilation, the use of modern editions for the compilation of corpora also results in various kinds of problems for linguistic research which have been demonstrated in a number of recent studies, including Kytö and Walker (2003), Bailey (2004), Grund, Kytö and Rissanen (2004), Lass (2004b) and Grund (2006). Perhaps the most extreme criticism has come from Lass (2004b), who sees edited texts as one of the most dangerous sources of corruption in the “forensic cleanliness” of historical corpora, because “of the degree to which they are trusted and characteristically regarded as ‘data’, worthy of inclusion in historical corpora” (22). In their description of some of the problems of traditional historical language corpora Honkapohja, Kaislaniemi and Marttila (2009: 456-60) list several which are related to the use of printed editions not originally designed for linguistic use as sources instead of original documents:

- 1) *use of critical editions* which conflate linguistic forms from various time periods, dialects and scribal traditions;
- 2) *varying editorial principles and loss of manuscript features* that either distorts analysis results or drastically diminishes the features that can be successfully analysed;
- 3) *predetermined research focus*, resulting from the preference of editors for literarily or culturally ‘significant’ texts which results in edition-based corpora containing a skewed selection of different text types;
- 4) *questionable orthography* resulting from frequent regularization of spelling in editions, which precludes many kinds of research;

- 5) *copyright issues* caused by the use of modern editions, which are still under the protection of copyright and prevent the corpus from being freely published or distributed;
- 6) *duplication of effort*, resulting from editions being published in a form that cannot directly be used as corpus text or automatically converted to the required format;
- 7) *shallow representation of manuscript reality*, with attention limited just to the level of the 'text itself', ignoring its visual, structural and paratextual aspects and thus precluding many kinds of pragmatic and discourse-analytical research.

Perhaps the most serious problem from a linguistic point of view is the use of critical editions that combine elements from several textual variants and thus intermix potentially widely differing dialectal and historical features or scribal practices. This creates a situation of linguistic hybridity representing the predilections of the editors instead of the language of the historical context from which the text supposedly originated (Honkapohja, Kaislaniemi and Marttila 2009: 456). An extreme case of this problem occurs with editions that employ conjectural emendation, where the edited text contains not only readings from different periods and dialects, but also purely editorial readings. At its worst, the use of editorially emended textual evidence—itself based on a view of what the text *should* have looked like—for making conclusions about historical language states can result in dangerous circular reasoning, as is demonstrated by the cases of the 19th-century editions of the *Life of St. Alexis* by Gaston Paris and of Henry Machyn's 15<sup>th</sup>-century diary or 'day book' by John Gough Nichols, reported by Cerquiglini (1989) and Fleischman (2000), and Bailey (2004), respectively.<sup>53</sup>

A related problem that affects the usability of editions designed for some other purpose as linguistic research material is that of orthographic normalization, which is often practised to a greater or lesser degree in both critical and documentary editions, at least to the degree of expanding abbreviations. This makes any corpus incorporating such editions unsuitable for the study of orthography or any other research question dependent on original spelling, as was pointed out by the compilers of the *Corpus of Early English Correspondence*:

Particularly the older editions (ie. the ones included in the CEECS) cannot be relied upon in questions of spelling, as the editors' priorities were often not linguistic but historical. Even [...] newer editions [...] [may be] a less than reliable source for studies of orthography.  
(Nurmi 1999: 55)

<sup>53</sup> Gaston Paris saw the linguistic irregularity of the *Saint Alexis* manuscripts as a result of scribal corruption and the influence of Anglo-Norman, and felt justified in eliminating this irregularity and 're-establishing' a regular system of French declension which he believed to have been in existence when the poem was composed (Cerquiglini 1999: 61-2). Subsequently, "these editions, and others of the same philological bias, have figured prominently in constituting the data base for grammars of Old French" and have been used by later editors to justify grammatical emendations in Old French texts (Fleischman 2000: 38). Henry Machyn's diary or 'day book' on the other hand was edited by John Gough Nichols in 1848, and despite serious infidelities in the orthography—which were only discovered by Axel Wijk in the 1930s—it was used as data for Wyld's *A History of Modern Colloquial English* in 1920 and 1936, and subsequently even sampled into *The Helsinki Corpus of English Texts* (Bailey 2004: 220-2).

The expansion of abbreviations, especially if performed ‘silently’, i.e. without indicating the originally abbreviated status of the word, is a problematic editorial convention, as it imposes editorial standardization on the text and obscures their original abbreviated orthography, which can be highly significant for many branches of linguistic study.<sup>54</sup> The expansion of abbreviations becomes especially problematic when the text of a printed edition is included in a corpus, as the traditional way of indicating expanded abbreviations, i.e. italicization, commonly gets omitted along with other typographical features of the text.

Although this problem is most acute in relation to the textual content of the editions, it occurs in a much more radical way in the case of *visual paratext*, which is very often either heavily normalized or even completely omitted and cannot thus be represented in a corpus based on the edition. While this might not seem to be a significant concern for many corpus linguists, the increased popularity of corpus-based pragmatics and discourse analysis means that there is increasing demand for corpora that also encode such “visual cues to levels of text structure” (Carroll 2006: 321-2) as layout, typography, rubrication and decoration (see subsection 4.3.3). Furthermore, as Pierazzo and Stokes (2010: 418) point out, it is very often impossible to accurately represent the textual structure of texts with multiple parallel streams of text—resulting from scribal emendations, additions and notes (see subsection 4.3.2 below)—without also reproducing the layout of the page at least to some degree.

A more insidious problem with using editions for corpus compilation is caused by the differing editorial principles of editions. It is not limited to critical editions, but also affects documentary editions, since they also vary in their principles of representing the original document (cf. *Practical methods of documentary editing* in subsection 3.2.1). The problem is exacerbated by the lack of clear documentation of editorial principles, which afflicts especially older editions (often preferred by corpus compilers for copyright reasons) and makes it impossible to evaluate the extent of the problem and to take mitigating actions. This problem was acknowledged already by the compilers of the *Helsinki Corpus of Historical Texts*, as pointed out in the Kytö (HC Manual): “editorial and typographical conventions vary in different source texts (e.g. emendations can be indicated by italics, parentheses, brackets etc.)”. The most common solution to this variability is to do as was done in the *Helsinki Corpus*: use a number of ‘text level’ codes to transfer the function of the convention to the “computerised version, irrespective of the particular format followed in the source text” (HC Manual).

However, while this practice might seem to produce a uniform result, it is in fact an illusion: due to the great variation in the degree of editorial intervention and in the extent to which different features have been annotated, the texts in the corpus either end up having a highly variable amount of details encoded in them, or must limit their annotation to those features which are annotated to a similar degree in all of the editions (Honkapohja, Kaislaniemi and Marttila 2009: 456). In either case, textual or paratextual features that are not recorded in all of the source editions, cannot be used for analysis. Furthermore, when critical editions exhibiting the problems of conflation mentioned above are used to build corpora,

<sup>54</sup> For example Schendl and Wright (2011: 25) and Trotter (2011: 155) note that abbreviations play an important role in neutralising the language of a lexical item and are thus important in identifying instances of code-switching and in deciding what language it should be attributed to.



their linguistic problems become compounded and obscured. In the best case, the limitations of individual editions are clearly documented and acknowledged in the corpus metadata; in a more likely scenario, the inclusion of the text in a corpus merely strips away what editorial documentation there was and obscures the textual nature of the edited text as well as any linguistic shortcomings (Honkapohja, Kaislaniemi and Marttila 2009: 456).

An entirely different kind of problem, having to do with the issues of balance and representativeness, is caused by the tendency of editors to focus on works that are considered to be somehow culturally or historically significant. The most obvious form of this bias is the preponderance of literary texts among editions of historical works, but in general, more ‘well-known’ works representing established genres tend to be more readily available in editions, which can easily lead to an unequal and historically skewed representation of different text types in a corpus (Curzan and Palmer 2006: 28).

### Need for linguistic corpus editions

All of the problems mentioned above are caused by the fact that the vast majority of scholarly editions are prepared for purposes very different from corpus linguistic research. Since for example literary historians see written documents of the past as “works of verbal art” (Wenzel 1986: 11) rather than witnesses to past linguistic acts, even digital editions produced for their needs are not likely to satisfy the needs of corpus linguists. Thus, as Dollinger (2004: 5–6) points out, it is actually somewhat unreasonable for corpus compilers to blame editors for not providing the kinds of source texts they would need. The fact that corpora based on edited texts will ever be “only as good as the editions and their editorial principles” (Dollinger 2004: 6) should motivate corpus linguists to develop editorial theory towards a more linguistically aware direction. Thus, in addition to merely acknowledging the impact of editors on the texts within their corpora, as Curzan and Palmer (2006: 26) urge, corpus linguists should also work with editors to bring about editions where the impact of the editor would actually be beneficial—or at least minimally harmful—to corpus linguistics.<sup>55</sup> As Lass (2004b) in his rather scathing critique of current corpus linguistic practice argues, minimally this would entail a digital edition that does not contain any editorial intervention that results in substituting the scribal text with a modern equivalent.

The production of digital editions that fulfil the needs of corpus linguists without the need for significant editorial activity by the corpus compiler would also help to eliminate duplication of effort between the production of printed editions and their subsequent digitization and compilation into corpora. This thesis and the present edition are based on the premise that taking the requirements of corpus linguistic research into account in the design and production of a digital edition

<sup>55</sup> Even editions prepared primarily for the purposes of literary criticism or historical research do not necessarily need to be incompatible with corpus linguistic research. Considering for example that the function of literary criticism “at its best” comes quite close to what can be seen as the function of pragmatic text linguistics, i.e. “to understand what texts are, how texts work, what texts have meant and can mean, and to understand these things within the contexts of language, history, value, politics, and the conflicting interests of the past and present” (Shillingsburg 1991: 41), it would not take much to make a digital edition prepared from such a point of view compatible with the requirements of historical corpus linguistics.

does not require all that much additional work from the editor, while both saving the corpus compiler a significant amount of work and resulting in a corpus/edition with improved analytical capabilities. Thus, one of the principal aims of the present edition and its documentation is to develop a model for a digital edition of historical documents that “would be immediately usable as corpus texts without a significant amount of additional work” (Honkapohja, Kaislaniemi and Marttila 2009: 459). The elimination of the discontinuity between edition and corpus will also mean that all of the textual and codicological features encoded in the edition are automatically available in the corpus. This enables the corpus compiler to rely on the editors for those aspects of the text that fall into their area of expertise (textual scholarship) and focus her attention on the analysis and linguistic annotation of the data (Honkapohja, Kaislaniemi and Marttila 2009: 462-3).<sup>56</sup>

### 4.3.2 Manuscripts as historical linguistic witnesses

Many textual disciplines whose theoretical frameworks have originally been developed in the context of modern printed texts, including literary and textual criticism and corpus linguistics, have not always been entirely successful in accounting for the peculiarities of manuscript texts. For example Markus (1997), arguing from a corpus linguistic point of view, sees scribal features such as “subpuncted, underlined and crossed out passages of text” as “specific problems of individual manuscripts which simply cannot be taken into consideration in corpus compilation” (218), despite the fact that they are practically ubiquitous in vernacular medieval manuscripts and in essence result in a nonlinear or a ‘multi-state’ text. In a similar vein, Sutherland (2009), coming from a background of 19<sup>th</sup>-century literature, claims that a ‘stable text’ is a requirement of most reading and scholarly purposes: “It is on notions of stable textual identity, persisting as shared cultural property, that reading communities are built” (22). From the point of view of medieval scholarship and manuscript editing, her insistence on the necessity of a stable text seems almost incomprehensible—for most medieval texts, there is and never was a ‘stable text’, and from the point of view of a historical linguist, producing one using the methods of traditional textual criticism would amount to falsifying the evidence.

Thus, if we wish to produce an edition ‘faithful’ to the ‘text’, we must first define what constitutes the ‘text’ that is the object of faithfulness. As the case of Duggan’s (1996) denunciation of modern conservative editing and Lass’s (2004b) subsequent critique of it demonstrates, the same edition can simultaneously be seen as extremely faithful and severely misrepresentative, depending on whether one takes as a reference point the linguistic act of the scribe or the intention of the original author. The disagreement between Duggan and Lass illustrates the importance of the *skopos* of both the edition and the text to be edited in selecting the editorial method. From the point of view of Duggan, whose focus is on the metrical choices of the original poet, an edition faithful to the scribal language of the surviving manuscript is just as useless as a text emended on metrical grounds is to a historical linguist like Lass, interested in the authentic language act produced

<sup>56</sup> Furthermore, all linguistic metadata that is added by the corpus compiler can also be turned into a new *annotation overlay* (see subsection 5.6.4) for the original edition and made available to its users, whether they are using the edition as a part of the corpus or not.

by the actual person copying the text.<sup>57</sup>

While Duggan is entirely correct in stating that “scribal spellings often misrepresent what the poet wrote” (228), it is equally true that as users of a particular historical language form and readers of the poet’s text they produced something that they apparently considered to be an acceptable rendering of the original text. Thus, manuscript versions dismissed by critical editors as worthless in terms of recovering the ‘authorial version’ of a text are often the most interesting in terms of contemporary intellectual engagement of copyist-editors, whose “activities provide a wealth of insight into contemporary or near-contemporary reading of a text” (Pearsall 1985: 103). This also means that from a linguistic point of view, each surviving manuscript version of a work is valuable, as different versions of a work do not “provide equivalent information, but information reflective of different historical and potentially historicisable situations” (Hanna 1992: 122). Thus, where a literary scholar sees a single work represented by multiple more or less corrupt versions to be used for the reconstruction of the artistic creation of the author, a linguist sees multiple records of linguistic events with a similar information content but with potentially different situational and even cultural contexts. This means that for a variationist corpus linguist interested in the effects of the external context on the linguistic features of the text, it is the *variation* between these versions that is the primary point of interest. While the close similarity of the versions in terms of their content helps eliminate a number of variables and focus on micro-level variation between the versions, it also causes problems in terms of their incorporation into a larger general purpose corpus, as will be pointed out below.

### Layered language

In addition to medieval works being represented by multiple textual versions, the mechanisms of their textual transmission mean that each manuscript version actually contains several layers of language use, representing not only the linguistic practice of the copyist responsible for that version but also of his or her predecessors. Scribes could treat exemplars written in a dialect different from their own in a variety of ways, accommodating them to their own linguistic practice to varying degrees and with varying consistency, while occasionally copying exemplar forms that are not native to their own dialect but rather to that of the exemplar.

This linguistic hybridity of documents containing various kinds of *relict* forms copied over from their exemplars has led to the question of “*whose* usage is actually reflected in the manuscripts and to what extent all the data in the manuscripts are valid for historical linguists” (Grund 2006: 116). As many historical linguists (see e.g. Benskin and Laing 1981, LALME, Laing 1989 and Lass 2004a) point out, this means that internal variation in these kinds of layered texts does not necessarily reflect variation in the usage of a single scribe, but may rather reflect layers of uses by several individuals. Although Grund (2006) sees the mixed linguistic

<sup>57</sup> In accusing modern conservative editors of being “by temperament and training committed to naive empiricism” and confusing “the document with the poem”, Duggan (1996: 223) in fact falls prey to the very same narrow-sightedness. What he fails to see is that it is not a question of confusion: those editors are simply more interested in the actual document than in what might have been the original authorial version of the poem.

usage entailed by the layered nature of medieval manuscript texts as a “facet of the well-known problem of ‘bad data’ in historical linguistics” (117), it does not necessarily need to be seen as a problem. Although the scribe responsible for a manuscript version of a text may not have originated or spontaneously produced its linguistic forms as a part of his or her usage, the fact remains that he or she still *accepted* them as sufficiently valid forms to be copied without changing them to something more appropriate. Thus, while the manuscript forms might not represent the active linguistic usage of the particular scribe copying the surviving document—who in most cases remains effectively anonymous in any case—they can be argued to represent his passive repertoire, linguistic forms considered understandable and acceptable within his discourse community (LALME: 1:14). In the case of medieval texts, this may well be the level of specificity that we have to settle for. This should by no means be taken as a deterrent to historical linguistic study, but rather as a reminder that “what we are studying is a linguistic feature as it appeared in a text over a period of time, and not one person’s employment of it” (Grund 2006: 117).

### Multiple versions and corpora

In order to build up as complete a picture as possible of the historical and dialectal development of *le langue* Middle English, we need all the information we can get from every Middle English text on the level of *la parole*, that is every Middle English scribal version of every Middle English text. (Lucas 1998: 171)

As mentioned above, the multiplicity and *variance* of medieval textuality—both in the form of multiple manuscript versions and ambiguous or variant readings within a single manuscript—is somewhat problematic for the traditional methods of corpus compilation and use, developed originally in the context of modern texts. As Curzan and Palmer (2006) point out, corpus compilers have traditionally equated a single corpus text with a single work, with the result that “a corpus (if built along the prevailing trends in corpus design) can represent only one choice among variants in different manuscripts or possible readings of a minim sequence” (26). This problem is partially a problem of current corpus compilation practices, and partially a problem in our predominant text ontology, which basically sees text as a “data stream” or a “a linear string of characters” (Sinclair 2004: 4) with no possibility for parallelism or divergence. The first of these problems, the question of multiple versions of a text and how they should be represented in corpora, has been discussed e.g. by Grund (2006) and is mostly a problem of balance, representativeness and statistical methods. The second part of the problem, however, could be seen as a more fundamental one, challenging our very notion of what text is.

On the level of text ontology, hypertext and other forms of nonlinear and variable text are nothing new, and even the technical issues of representing such a text in machine-readable form have been quite satisfactorily solved.<sup>58</sup> The aspect

<sup>58</sup> For example the Text Encoding Initiative’s *Guidelines for Electronic Text Encoding and Interchange* (TEI Consortium 2014) provide the means to encode parallel (either concurrent or mutually exclusive) passages of text as well as the editor’s assessment of their relationship on several levels of complexity (e.g. 78–84, 351–84, 513–21, 531–6).

of corpus linguistic methodology that is much less attuned to the peculiarities of these kinds of multilinear or indeterminate texts is their textual and linguistic analysis. All of the commonest tools and analysis methods used by corpus linguists are based on the linear character string model of text, and are thus ill-equipped to deal with variant, indeterminate and multilinear texts, which would require a more probabilistic approach acknowledging the simultaneous existence of multiple textual states within a single textual object.

In the case of works existing in multiple versions, a traditional critical or eclectic edition is an anathema for the variationist approach, since it completely ignores the variation between versions of the work. Like Grund and Machan (1994: 190-1), this thesis considers the ideal situation to be a digital edition containing transcriptions of all the available versions (Grund 2006: 118) that enables researchers “to capture variation across different manuscript copies” and study the “scope and characteristics of any given case of morpho-syntactic variation” within the textual tradition (119). While this kind of a digital edition can accurately represent the variation inherent in the textual tradition in a computer-readable format, it does not solve the problems inherent in including such a text in a corpus (Grund 2006: 121-2).

Grund (2006: 119-21) sees two basic ways of approaching the problem, both of which have their benefits and drawbacks. The first approach is to treat each manuscript version as an independent corpus text, which he sees as a reasonable option in the case of texts with significant differences between the versions. This approach would allow access to the full variation between the manuscript versions, and would also seem to be the approach envisioned by Lass (2004b), since he uses as his model the *Linguistic Atlas of Early Middle English (LAEME)* corpus, which contains several manuscript versions of a single work. While this first approach is the natural one e.g. for dialectology (which is Lass’s focus), it is problematic for quantitative morpho-syntactic studies, as having multiple versions of the ‘same text’ in a corpus would mean, first of all, that all those parts of the text where there is no variation would occur several times and skew quantitative studies, and second, that the data from a multi-version text will be overrepresented in comparison to data from a text surviving in only one version (120).

Grund does, however, point out that “if we take the extreme position” and view each text as an independent linguistic object or *utterance*, we can argue that this does not matter as each scribe “can be claimed to have had the choice of keeping a feature or replacing it with another” (120).<sup>59</sup> The second approach identified by Grund to the problem of multi-version texts in corpora is to consider all manuscript versions of a multi-version text as manifestations of the same text and include only one version of the text in the corpus. This approach also has its share of problems. First of all, there is very rarely an obvious ‘best text’ candidate that would be somehow ‘more representative’ of the text than the other versions. One could also use extralinguistic criteria and choose e.g. a version whose scribe is known, but in the absence of such criteria (as is the case for most scientific and practical manuscript texts), the choice would be purely arbitrary. Including only one version of the text also carries some of the same problems as using a criti-

<sup>59</sup> He does, however, also point out that this assumption is in fact an oversimplification, since “it is questionable whether all linguistic uses in a manuscript, especially morpho-syntactic ones, can be considered potentially substitutable”.

cal edition, as all inter-manuscript variation is lost, along with a potentially large number of linguistic features that do not happen to occur in the selected version but would occur in the omitted ones.<sup>60</sup>

For an optimal compromise solution for including multi-version texts in corpora, (Grund 2006: 121) proposes the inclusion of all the different manuscript versions, combined with the ability of the user to choose one of the versions as a ‘primary’ one for quantitative analyses, where the inclusion of several versions would be clearly problematic. This is also the approach adopted for the present edition. The selection of a ‘primary’ manuscript by the user is supported by comprehensive metadata about each of the manuscript versions both in structured format in a header included in the source file of each transcription (see section 11.1) and as descriptive prose in section 9.2, as well as a description of the textual relationships between the different manuscript versions in chapter 13. While this solution places all of the data at the user’s disposal, these kinds of multi-version texts still pose a challenge for current quantitative corpus linguistic methods.<sup>61</sup>

### Implications for editions and corpora

Instead of approaching medieval manuscripts merely as containers of stable, linear and self-identical texts, we should see them as documentary evidence in a very concrete sense: as a physical remnant of a historical linguistic event. Lass (2004b: 21) quite aptly likens the historical linguist’s workspace to that of an archaeologist or a forensic investigator, pointing out that it is a commonplace in both of these disciplines that “as far as possible, (a) the scene must not be contaminated by material brought in by the investigator or anyone else, and (b) the chain of custody (the sequence of provenances of all objects found on the scene) must be immaculate”, maintaining that the same should be true of “any linguistic historiography” (Lass 2004b: 23). Since manuscript texts can be considered to represent for a historical linguist “*the utterances of our informants in a fieldwork situation*”, and “good field-workers do not gussy up their field-notes *qua* field-notes, though they may ‘normalize’ for some purposes later”, historical linguists “do not have the right to normalize for any purpose, or to publish or use as ‘materials for the history of English’ the work of editors who have normalized or emended” (28):

Ce que le linguiste contemporain obtient de plus précieux du locuteur natif d’une langue vivante, à savoir le mouvement langagier, le jeu de la forme et du sens, tout cela est exposé par les manuscrits médiévaux, mais dispersé par leur édition.<sup>62</sup> (Cerquiglini 1989: 108)

<sup>60</sup> Based on his examination of personal pronouns used to refer to inanimate objects in the *Mirror of Lights*, Grund (2006: 113-5) observes that using only one or a few of the seven manuscripts of the text would have resulted in different statistical results, since there are major qualitative and quantitative differences in the use of these pronouns between the different manuscripts. Thus a qualitative analysis based on one manuscript could very well reach conclusions that would not hold if several or all of the copies had been consulted.

<sup>61</sup> As one solution for the quantitative analysis of multi-version texts as a part of a corpus, Grund (2006: 121) proposes the use of the mean frequency or aggregate total frequency of a linguistic feature (depending on the research question) in the different manuscript versions as the value representing the whole in order to make the multi-version text quantitatively compatible with single-version texts.

<sup>62</sup> “The most precious thing that the contemporary linguist can get from the native speaker of a living language, that is, the dynamic of the language, its play of form and meaning, is all there to be seen

It is important that an edition or a corpus represents the multilinear and variant nature of medieval manuscripts both on the level of variant manuscript versions and of variation within individual manuscripts, separating from each other the contributions of different scribes and preserving both the initial and final versions of passages emended either by the original scribe or later correctors. This allows linguistic features discovered in the text to be assigned to the scribe responsible for them and their relative chronology and mutual relationships to be determined. However, what we still do not have are technical tools and quantitative analysis methods that would allow us to analyse these kinds of multi-version texts, or as theoretical models that would allow us to relate them to the big picture of language use.<sup>63</sup> Despite these wants in our tool kit—which will hopefully turn out to be temporary and are already circumventable with the creative use of standard programming technologies—this kind of variation and parallelism *does* exist in our data and we need to either take it into account, or at least explicitly acknowledge our failure to do so.

### 4.3.3 Practical requirements of corpus linguistics

I begin with the fact that editions are composed to communicate texts, and the fundamental choice of editorial technique is not based on some property inherent in the materials edited but on the communication which the editor intends to effect. That is, editorial technique isn't God-diven or absolute but a fact about audience or about critical perspective. (Hanna 1987: 87-88)

As the above quotation points out, there are no generally appropriate editorial methods, but their selection is now generally acknowledged to be dictated by the intended use of the edition (see e.g. Machan 1994, Rehbein 2008 and Robinson 2009). As has been already established, the editorial model described in this thesis and exemplified by the present edition is designed to serve as research material for historical corpus linguistic research and targeted at corpus linguists who wish to not only perform linguistic analyses but also encode their results in the form of new *annotation overlays* (see section 5.4) to be shared with other scholars.

As was pointed out in Honkapohja, Kaislaniemi and Marttila (2009: 2), the field of historical linguistics has over the last decade seen some discussion about the specific requirements for digital editions suitable for historical linguistic study (see for example Bailey 2004, Dollinger 2004, Curzan and Palmer 2006 and Grund 2006). While the practical needs of corpus linguists are not especially difficult to fulfil, they are not only quite different but also partially contrary to those of literary scholars and historians, who are the traditional audiences of scholarly editions. Whereas literary scholars are interested in the artistic work embodied by a document and its interpretation, and historians in the historical significance of its content, historical linguists are interested in the linguistic forms contained in the

---

in medieval manuscripts, but then it is dissipated when they are edited.”(Cerquiglini 1999: 75)

<sup>63</sup> The online user interface and toolkit for the present edition and future editions following the guidelines presented in this thesis that is planned as a separate project, will incorporate basic corpus linguistic search analysis tools that will allow both the selection of a ‘primary’ version to represent a multi-version work and the calculation of various kinds of mean and aggregate values for multiple versions.

original document. Furthermore, linguists—like historians—are interested in documentary texts, i.e. texts as they occur in a single document, but unlike historians and literary scholars, they are usually interested in not only the ‘substantial’ features of the text, but also those that are usually considered ‘accidental’, including things like orthography, punctuation, and even such paratextual features as layout and typography. In addition to differences in the features required of an edition, linguists are also interested in different *kinds* of texts than literary scholars or historians, tending to emphasize the typical and everyday over the exceptional or creative.

In addition to these requirements and preferences concerning the kinds of sources suitable for corpus linguistic editing, the principal practical requirement is that the edition models the content—however defined—of the document in machine-readable form. This requirement, although seemingly obvious in the context of digital editing, means that pure digital facsimile editions, as useful as they might be for many other purposes, are virtually useless from the point of view of corpus linguistic methodology.<sup>64</sup> Since the inclusion of facsimile images in an edition, while obviously desirable, does not relieve the editor of any significant amount of work—although they often do make this work much easier—the inclusion of images is often not worth the significant fees many libraries still charge for providing images of their holdings and especially for a license for publishing them online.<sup>65</sup>

On a conceptual level, Lass (2004b) has formulated what he calls “three inviolable desiderata” that need to be fulfilled by “a proper historical corpus” (40) and by extension, by a digital edition that is eligible for inclusion in a historical corpus:

- 1) maximal information preservation,
- 2) no irreversible editorial intervention, and
- 3) maximal flexibility.

Lass (2004b) defines these three general principles on a more practical level in the negative, listing a set of features—most of which associated with critical editing but some also with the more ‘liberal’ forms of documentary editing—that should immediately disqualify a source from being admitted as a linguistic witness (22):

- 1) Any emendation, even of what appear to be patent errors.
- 2) Any modernization, including the replacement of thorn, edh, yogh and wynn by modern equivalents; capitalization practice different from that of the source; alteration of MS punctuation, whether by modernization of the original or punctuation of an unpointed text.

<sup>64</sup> This does not mean that digital images could not be an extremely desirable supplement for a corpus linguistic edition, providing the user additional details of the visual aspects of the document and allowing her to verify the editorial readings of the text, but they simply cannot be used as data for corpus linguistic research.

<sup>65</sup> Fortunately, the attitudes of libraries and especially archives are slowly changing, and more and more libraries and archives are either digitising their collections for open online publication or allowing editors to take their own photographs of the material. Unfortunately, the only images currently available for the documents edited in here are black and white microfilm images which cannot be freely published without substantial fees. This means that the linking of images to the edition must by necessity remain a project for the hypothetical future until better-quality images are made freely available by the respective holding libraries.



- 3) Any alteration of scribal word-division or lineation, including the printing of apparent verse texts written as continuous prose in verse form.
- 4) Any attempt to reconstruct a 'lost original' or 'archetype' from a multi-source tradition, or to produce a 'best text'; in other words any multi-sourced or conflate reading-text, such as the standard editions of Chaucer or Shakespeare.
- 5) Any form of 'normalization', e.g. regularising variable spellings of a given lexeme or grammatical form, or dialect translation.

Johansson (2004: 97), discussing the requirements for a digitally edited text to be included in a *text archive* of original source materials—a kind of a historical text corpus—lists three major considerations to be taken into account:

- 1) *Accessibility*: the archive should make it easier for its users to gain access to information and materials that would otherwise be available only in different libraries or archives;
- 2) *Presentation*: the archive should provide a visual representation of the original texts on the computer screen, limiting the need to use the original; and
- 3) *Searchability*: the texts in the archive should be searchable for their textual content and preferably also for other features.

While the emphasis would be slightly different for linguistic corpora, these general principles are equally applicable for them. While Johansson sees the obvious solution for the first two considerations to be the inclusion of facsimile images of the original, this solution fails to answer the third requirement, which necessitates the text to be transcribed and preferably also annotated in some way. Instead of a facsimile—which should be considered an additional luxury rather than a necessary feature in a linguistic edition—the method of fulfilling the second requirement advocated by the present thesis and used in the present edition is the detailed descriptive annotation of the *visual paratext* integrated into the transcription (see below). Based on the observations of earlier scholars and my own practical experience in both digital editing and corpus linguistic research, a digital edition intended for historical corpus linguistic research should exhibit the following qualities and features:

- 1) searchable diplomatic representation of textual content,
- 2) representation of visual paratext and text structure,
- 3) explicit indication of responsibility,
- 4) flexible access to source data,
- 5) integrability of research results, and
- 6) reliable documentation of sources and methods.

### Searchable diplomatic representation of textual content

My hope is that we, as a generation of scholars, recognize that, in order to apply the new technology of the electronic text most profitably for our own and succeeding generations, our first duty is a peculiarly self-effacing one. We should now go through a period in which we produce a huge number of simple, accurate, faithful editions, with the minimum of editorial intervention of whatever sort.

The objective is not to preside over the text, explaining and controlling it like the magisterial anatomist of the sixteenth century, but to act as its midwife, ushering it forth to let it speak for itself in good time.  
(Leslie 1993: 50-51)

In terms of historical corpus linguistics, the most important requirement for a digital edition<sup>66</sup> is the faithful transcription of the textual content of a single document. From a linguistic point of view, the definition of ‘faithful representation’ is here considered to be accurate *graphemic* transcription. The formulation *textual content of a document* is here used consciously instead of the simpler *text* to acknowledge the fact that each medieval manuscript contains more than one text, and it is important that the transcription accurately reflects the layered and variant nature of these texts by recording all alternative readings, such as the original and corrected form of a passage emended by a scribe or a later annotator, and annotates their parallel status explicitly. The transcription should also include such orthographic features as punctuation and abbreviated forms, as well as any pictorial or symbolic elements occurring in the document. While it is impossible to produce a perfectly accurate and objective digital transcription of a manuscript text, “linguists transcribing for linguistic research” are nevertheless likely to “render us with more reliable data in our corpora” than would relying on modern editions produced for non-linguistic purposes (Dollinger 2004: 11-2).

As obvious a choice as faithful graphemic transcription seems from a linguistic point of view, it is in fact quite far from the editorial mainstream, as emendation and normalization (of punctuation, layout and even “‘unusual’ spellings of familiar words” (Lucas 1998: 174)) is taken for granted by many traditional textual scholars (e.g. Edwards 1998a and Blake 1998) and practised even by documentary editors. Even though F. W. Maitland recognized the importance of preserving original spellings already in 1903 (Marvin 2004: 22),<sup>67</sup> even linguistically oriented editors and corpus compilers like Markus (2000: 182) have considered it perfectly acceptable—or even necessary (Markus 1997: 212-3)—to modernize the punctuation and capitalization of texts.<sup>68</sup> The practical question of normalization clearly continues to be a difficult issue for corpus linguists. As recently as 2008 both Hunston (2008: 158) and Claridge (2008: 252) have described the question whether to normalize unconventional spellings as an important decision for corpus compilers. While the regularization of spelling can help in automated linguistic analysis,

<sup>66</sup> Since we are here discussing exclusively *digital* editions, the requirement of machine-readability, which is the foundation of modern corpus linguistics is taken as a given.

<sup>67</sup> Although he did modernize things like capitalization and punctuation and expanded abbreviations, his argument for preserving original orthography was based both on the inherent interest of non-standard forms and on the observation that if we start to ‘correct’ the text, it is very difficult to tell where we need to stop.

<sup>68</sup> A good example of this is Markus’s (2000) praise of the editors of the Western Michigan University Middle English Text series, who “claim that their editions ‘maintain the linguistic integrity of the original work but within the parameters of modern reading conventions’” (182) by printing them in modern alphabet and following modern punctuation and capitalization practices. For example Lass (2004b) has quite strongly disagreed with this view, arguing that the “practices Markus praises (and which he carries to extremes in his programme of ‘normalization’) display quite the opposite of scholarly ‘integrity’: they produce falsifications, things that never were, given the names (or the invented names) of things that still are, if only they were accessible” (Lass 2004b: 33).

it also means that corpora based on editions exhibiting any degree of normalization cannot be used for the study of orthography, dialectal variation or any other research question dependent on original spelling.

Fortunately, the problems that Markus (1997: 212-3) has listed as forcing editors of Middle English texts to practice normalization of the manuscript text, namely “fonts and letters; punctuation and diacritical signs; abbreviations; format and layout”<sup>69</sup> have been solved by the development of the Unicode standard and markup technology (SGML and XML) and guidelines (TEI)—all of which existed already in 1997, but had obviously not yet penetrated the corpus linguistic community. With regard to the issue of normalization, this thesis agrees with Lass (2004b: 36) that while the normalization of orthography can be very useful for certain kinds of linguistic research questions, normalized transcriptions should not be considered primary data for historical linguistics and used as such in corpora. Rather, I argue that normalized or regularized word forms should be seen as a useful form of *analytical annotation* (see section 5.6) that can be linked to a faithful diplomatic transcription to provide the scholar’s *interpretation* of the lexical identity of word units and help in word-sense disambiguation.

### Representation of visual paratext and text structure

In addition to the accurate transcription of the textual content of a document, some scholars (see e.g. Grazia and Stallybrass 1993, Caie 2003, Meurman-Solin 2007, Del Lungo Camiciotti 2011, Smith and Kay 2011, Suhr 2011, and Tyrkkö, Marttila and Suhr 2013) have over the past two decades adopted the New Philological position and called for “a broader kind of bibliographic approach, which does not see text as essential and physical features as dispensable” (Eggert 2004: 164), but rather sees the entire physical document as a vehicle for communication that contains “data of considerable interest and value that cannot be elegantly or effectively communicated by transcription alone” (Leslie 1993: 46). For example rubrication can indicate subdivision of the text into chapters or sections, and the layout of text can provide clues about its nature and intended use (Edwards 1998b: 164). This means that omitting the layout and physical structure of the text and normalising the textual structure of the text—which is a common practice for editions of literary texts—can lead to significant misinterpretations and misunderstandings about the nature of the text (Cooper 1998: 85-6).

As has already been pointed out on several occasions, this thesis very much shares this view and thus argues emphatically for the importance of editing not the text but rather the entire document, as far as this is possible. In practice this means that in addition to the textual content, the digital edition should also extend the same principle of faithful representation to the visual paratext of the document—including its layout, typography and decoration or highlighting—on a level that is parallel to the text (see sections 10.1, 11.3, 11.4 and 11.5). Although this kind of *descriptive annotation* (see section 5.6) can only represent some selected visual aspects of the original, its abstract and analytical nature not only allows the reasonably accurate visual representation of these features, but also allows them to

<sup>69</sup> Markus’s use of the term *font* in this context is very confusing, as later on he does discuss the different types (italic, boldface) used by editions to indicate manuscript features, but here he seems to be referring to the use of different scripts in the original manuscripts.

be used as search criteria and factors in quantitative analyses along with the text and analytically annotated text-structure (see sections 10.1 and 11.2).<sup>70</sup>

As noted above, corpus linguistic methodology is based on the discovery of correlations or dependencies between linguistic and extralinguistic features. In order to be able to relate the linguistic features of the text to their *textual context* in addition to their physical (documentary) context, a digital edition prepared for the needs of corpus linguists should also explicitly encode the textual or logical structure of the text. In practice this means dividing the textual content of the document into structural units (parts, chapters, sections and paragraphs, lists, headings, etc.), allowing the status of each segment of text to be evaluated with respect to the rest of the text. In the present edition this kind of elementary analytical annotation serves two functions. First of all it allows the corpus linguist to evaluate whether a certain feature occurs more frequently in one specific section of the text than in others, or whether the occurrence of a certain feature is limited only to certain type of textual component (e.g. headings), and secondly—and more importantly—it serves as the basis of a *textual coordinate system* to which other analytic annotation can be related. It also serves to delimit the size of not only the text, but its subcomponents, extending the important corpus linguistic requirement of finite and known size below the level of whole texts, making it easy to relate quantitative findings to e.g. the total number of words constituting headings in the text, or to the size of the paragraph in which a given feature occurs.

### Explicit indication of responsibility

As mentioned above, a digital edition prepared for the purposes of corpus linguistics needs to record not only a single editorially constructed linear text, but all of the textual content found in the document and to explicitly encode the status of any alternative or parallel readings (scribal emendation, glosses or annotations located parallel to the main text, etc.) in order to allow the reconstruction of all the possible texts encoded in the original document. From a corpus-linguistic point of view, however, this is not enough. As was mentioned above, variationist corpus linguistics is highly dependent on contextual information about the text, and considering that we know nothing of the *author* and *principal* (as defined in subsection 2.3.1) of most medieval texts—especially utilitarian ones—the *animator* of the text, i.e. the scribe, becomes the primary means of contextualising the text through attribution.<sup>71</sup> As Lass (2004b: 27) has pointed out, many kinds of linguistic analysis—especially dialectal and sociolinguistic ones—focus on the idiolect of an individual, which means that it is vitally important to know whether an addition or correction to a manuscript text or a marginal note was inserted by the

<sup>70</sup> In addition to providing additional data for analysis, the representation of the physical paratext can also help to mitigate the effects of what Rissanen (1989: 16-7) has called “the philologist’s dilemma”, namely that corpus linguistics could damage the discipline of philology by discouraging students from engaging with Middle English texts in their original form, by situating the abstracted results of corpus searches in their original documentary context and promoting a more ‘philologically responsible’ strand of historical corpus linguistics.

<sup>71</sup> Even if the scribe cannot be identified, as is the case with most manuscripts, the palaeographical features of the scribal hands found in the document can help in distinguishing and possibly even dating the various textual layers of the document.

original scribe or someone reading the text two centuries later. This means that a digital edition designed for historical corpus linguistics needs to document the scribal hand responsible for each segment of text in the document, including any additions and deletions made to it after its writing, since this allows us to link together all utterances produced by a single scribe and thus to construct a coherent linguistic profile and assign it to for a single language user and any biographical metadata—however scant—available about him or her.

The same naturally applies to any emendations, clarifications and modifications made to the text by the editor, such as the expansion of abbreviations or the correction of so-called ‘obvious mistakes’. For example, the transcription should always note the actual visual appearance of an ‘erroneous’ passage, an abbreviated word or a roman numeral, and encode the editor’s corrected version, interpretation of the abbreviation or the formal expression of the number in modern notation separately, explicitly indicating their editorial status. While printed editions reconstructing a single text on the basis of the document were forced to prioritize either the documentary or the editorial reading of a problematic (erroneous or unclear) passage by including it in the text and relegating the other to a note, the digital format allows the editor to encode both readings with equal priority, in effect providing her interpretations with more leeway, as they are no longer supplanting but rather expanding and commenting upon the original documentary readings (Lavagnino 1995: 119–20).<sup>72</sup> This means that editorial interpretations can be included without “intrusion of alien material into the historical record, or a loss of genuine material” (Lass 2004b: 22).

### Flexible access to source data

What prevents many digital editions from being used for linguistic research is the limited access they provide to their source data. As was pointed out in subsection 4.1.2, corpus linguistic research involves searching for linguistic patterns in the text, and then sorting and filtering—often manually—these search results in various ways to locate all instances of a given linguistic phenomenon in the text, followed by their analytical examination in relation to various contextual factors. Minimally, this requires for the textual content of the document to be available in a faithful, machine-readable transcription as described above. The concept of ‘availability’ in this context involves both a technological and a political aspect.

While the technical aspect can at first seem to be related exclusively to issues of user interface (see section 5.1), it is in fact intimately related to the issues of data collection and storage. First of all, the edition should contain a full transcription of the textual object and preserve its original textual organization, allowing its linguistic features to be studied in the context of its discourse structure. In the case of multi-version textual objects, all of the versions should be included,<sup>73</sup>

<sup>72</sup> As it has been the usual practice to prioritize the editorially emended reading by inserting it into the edited text and to relegate the original form to a note, corpora based on these kinds of editions have often ended up incorporating the editorial form in the corpus and stripping away the original one along with the editorial notes (Lass 2004b: 27).

<sup>73</sup> The virtue of extensibility provided by the digital medium means that the editing of a work occurring in a large number of manuscript versions does not need to be undertaken at once but can be divided between several separate projects employing a shared encoding standard and a set of editorial guidelines. However, the final aim of an editorial project should be the inclusion of all

their textual content linked between the different versions on a suitable level of granularity. Furthermore, both the textual transcription and the description and analysis of visual paratext and textual structure (described above) should be encoded in a *structured* manner and in recognized standard formats—such as the Unicode and eXtensible Markup Language (XML) used in the present edition—to ensure maximal compatibility with existing and future software technologies that are used to provide it with analysis and visualization tools. As was pointed out above, the implementation of a data structure defined by a standard interface would help ensure that the data can be accessed not only by proprietary tools developed for the specific edition, but by any tool or user interface that implemented that interface.<sup>74</sup>

The other aspect of accessibility is a question of editorial policy. Many digitization and digital editing projects produce “sophisticated, information-dense computer files” as their archival versions, but make available to the public only lighter derivative files that either omit valuable data or present it in a format that is not as well-suited for analysis, being often geared towards display (Dahlström 2009: 43). Instead, editorial projects aimed at corpus linguistic use should make available a maximum amount of information in a structured format; this allows researchers to decide what parts of the available information are relevant to their current research interests and filter out unnecessary information only when it is clear that it is not useful. Thus a digital edition intended for linguistic use (or in fact for any research use) should provide all of its data—here considered to include all transcriptions, any annotation linked to them, and any computer scripts produced for linking them together or for other kinds of manipulation and conversion of the data—under a license that allows not only their passive viewing but also their free processing for research purposes, i.e. use with third party analysis tools and aggregation with other text resources or other kinds of data.

### Interoperability and integrability of new data

In addition to allowing flexible access to its data, a digital edition designed as a corpus linguistic resource should also make use of the extendability provided by the digital medium by making it easy to link new data to the edition or to link data contained in the edition to other resources. Cummings (2009) has called these kinds of editions that facilitate interoperability with other resources through linking “agile interoperable editions” (307). Considering that the basic process of corpus compilation is the linking together of a principled selection of texts into a single resource, and that the manual analysis of concordance results that is often an integral part of corpus methodology results in a large amount of text-specific analytical metadata whose integration into the corpus would be very beneficial, this kind of agile interoperability should be a central requirement for a corpus linguistic edition. However, this need for the convenient linking of texts and other kinds of data is neither new nor unique for corpus linguistics. For example Sanders (1995) has argued that digital editions intended for any kind of research should

---

manuscript versions in order to provide as rich a dataset as possible for the study of linguistic variation.

<sup>74</sup> The specific features or properties required of such an interface are beyond the scope of the present edition and will be discussed separately in a later article.

allow the user to “gather and link texts for themselves”, since these operations are central to textual scholarship and the “quintessential mental operations necessary for complex creative use of our increasingly massive literary inheritance” (138).

On the level of general principles, technical interoperability is promoted by the employment of standardized data structures, encoding formats and editorial practices<sup>75</sup> and their detailed, comprehensive and unambiguous documentation. On a more practical level, an important precondition for the easy linking of text and data is the existence of a robust *coordinate system* that allows individual textual elements or pieces of data to be identified and addressed. Since corpus linguistic research operates on a finer level of granularity than most forms of textual criticism, dealing with individual word-units and even with individual morphemes, the text-structural annotation described above—usually extending only to the level of headings and paragraphs—is not fine-grained enough. In order to provide unambiguous and robust ‘textual coordinates’ to which linguistic and other analytical metadata can be anchored, each word-unit should be explicitly delimited and provided with a unique identifier.

In addition to allowing the editor to use it for conveniently linking external metadata—both editorially created and found in external resources—to individual words, this kind of a *textual coordinate system* also allows the users of the edition to create new metadata *overlays* (see subsection 5.6.4) linked to the base text on the level of individual textual units without having to make changes to the original edition. These kinds of overlays, being separate from the edition but persistently linked to it, can also be shared with other scholars and thus form the basis for the cumulative buildup of metadata. In the corpus linguistic workflow this means that all search results obtained from a corpus with such a coordinate system are persistently linked to the original texts, and the results of any manual filtering or analysis of these results can be stored as a separate metadata overlay and associated with the original edition, resulting in the gradual accumulation of metadata.<sup>76</sup> This concept of textual coordinates forms a central component in the present edition and the editorial approach advocated by this thesis. The practical implementation of this scheme in the present edition is described in subsections 10.1.3 and 11.2.5.

### Reliable documentation of sources and methods

We cannot study the history of English language without well-understood ideas about the reliability of the material that makes up the data for our history writing. (Bailey 2004: 217)

<sup>75</sup> Of these three, editorial practices are by far the most difficult to standardize and are therefore in need of most work.

<sup>76</sup> For example, supposing a corpus linguist doing specific research on a corpus built of such editions finds all instances of a certain lexical item in the corpus and analyses them for their syntactic functions. While this kind of data is traditionally discarded once it has been used to answer a specific research question, this kind of a textual coordinate system would allow the corpus linguist to export the annotated search results—including the word identifiers embedded into them—into a set of new metadata overlays, one for each edition in the corpus. These overlays, consisting of word identifiers and the syntactic functions associated with them, could then be exported into repositories associated with the different editions and there automatically merged with other overlays containing syntactic information on different lexical items.

The importance of documentation has already been mentioned several times in different contexts, and for good reason, as it is essential for achieving at least some semblance of reproducibility—one of the cornerstones of the scientific method. Although editorial processes are very rarely reproduced—and entire editions practically never—documentation of the source materials and editorial practices of an edition allows its users to understand and evaluate its nature, its possibilities, and its limitations. Documentation can in this context be divided into two components that differ somewhat in their nature and serve slightly different purposes: the documentation of sources and the documentation of editorial methods. Together, the documentation of both the original document and the practices and procedures followed by the editor in creating the edition establish the relationship between the original document and the edition as a *model* of that document (see chapter 5 and section 5.4).

The documentation of sources, i.e. the original documents that the edition is based upon, is important for two main reasons: it identifies the original document that the edition aims to represent, anchoring it to the world of physical phenomena, and characterizes the original document in various ways, providing contextual information to which the user of the edition can relate the textual and linguistic characteristics of the textual object. The production of documentary metadata, which provides information about the textual and intellectual content of the document, the properties of the manuscript as a physical object, and the known history and provenance of the manuscript belongs to the fields of codicology and descriptive bibliography.<sup>77</sup> In the context of corpus linguistics, texts are often also provided with analytical documentary metadata that provides information about the cultural and social identities of the textual object (register, genre, text type, target audience, etc.) and the people involved in its production (traditionally the author, but potentially also the scribe or printer, etc.), which help to situate the textual object into its situational and cultural contexts. In corpus linguistics, descriptive metadata about the textual object is also used for categorising texts based on their extralinguistic properties and to define subcorpora, for which reason it is important that the descriptive metadata is provided in a structured and machine-readable format.<sup>78</sup>

The documentation of the editorial method is here understood as the description of the procedures by which the edition has assumed its current shape, i.e. the mapping between the original document and the edition. If the documentation of sources provides the user with information on *what* the edition represents, the documentation of editorial methodology provides information about *how* it has been represented.<sup>79</sup> Since editing—whether digitally or in print—is always a form of reductive modelling (see chapter 5), the documentation of methodology needs to account for 1) *what* features of the original document have been represented in the edition, and 2) *how* they have been represented. This information is required

<sup>77</sup> This type of information is traditionally contained in library catalogues, although in practice the extent of the information provided by existing library catalogues varies widely.

<sup>78</sup> Documentation for the sources of the present edition is provided as prose description in section 9.2, and in structured form in the metadata headers found in the transcription files of each manuscript version.

<sup>79</sup> The documentation of editorial methodology should describe both the procedures and techniques used to transcribe and proofread the texts and the technical details of the character encoding and markup systems used to represent them in the edition (Ore 2004: 42).



not only for evaluating what kinds of research the edition can and cannot be used for, but for comparing analysis results between editions with differing editorial practices. While the documentation of editorial methodology should be a given for any scholarly edition, whether digital or printed, reliance on established and generally accepted practice has often led print-era scholarly editors into taking their transcriptional and editorial methods for granted and leaving them undocumented. The present edition and its documentation takes the opposite attitude and errs on the side of caution, preferring to document its editorial practices as completely as possible, even at the risk of occasionally stating the obvious. The editorial practices and annotation conventions of the edition are documented concisely within the metadata header of each transcription file of the edition and in more detail in chapters 10 and 11.<sup>80</sup>

#### 4.3.4 Earlier linguistic editions

In lamenting the unsuitability of most scholarly editions for corpus linguistic study, Lass (2004b) has claimed that the editions “that sell nowadays are generally for literary audiences, and the idea seems to be to produce that chimaera ‘the reading text’”, which means that “material for proper linguistic historiography becomes less available over time, unless special efforts are made” (28-9). The recognition of the unsuitability of traditional editions for corpus compilation has also affected the compilation principles of many recent and ongoing corpus projects, causing them to move increasingly toward using original documents as their sources.<sup>81</sup> One of the basic ideas of the present edition and this entire thesis is to combine the strengths of digital editions and linguistic corpora into a single multi-purpose research resource. Considering that these two kinds of digital resources are very similar in nature, there have so far been surprisingly few attempts to combine them (Honkapohja, Kaislaniemi and Marttila 2009: 461).

This is not to say that there are no predecessors to editions like the present one. Two examples of projects whose stated aims are very similar to the present edition—even if the practical solutions used to achieve are very different—include the *EWD* project and the compilation project of *ACOMESP*. In a similar vein to the present edition, the editors of the *EWD* proclaim that their edition “will be geared to facilitate advanced computer searches” and that they “combine [their] philological and editorial aims with principles of modern corpus compilation, striving at a new type of text edition that will also serve as a computerized corpus” (Kytö, Grund and Walker 2007: section 5). *ACOMESP*, on the other hand, offers a web interface that allows facsimiles (provided by the Hunter collection of the library

<sup>80</sup> These chapters will also form the basis for generalized prescriptive guidelines for producing digital editions of historical documents for the purposes of corpus linguistics, the formulation of which will form a separate postdoctoral project.

<sup>81</sup> Corpora that have either been compiled or are being compiled from original manuscript sources include the *English Witness Depositions 1560-1760: An Electronic Text Edition (EWD)*, compiled at the University of Uppsala; the *Middle English Grammar Corpus (MEG-C)*, compiled at the University of Stavanger; the *LAEME* corpus, compiled at the University of Edinburgh; the *Linguistic Atlas of Older Scots (LAOS)*, also compiled at the University of Edinburgh; the *Corpus of Early Ontario English* compiled at the University of British Columbia; *A Corpus of Middle English Scientific Prose (ACOMESP)* compiled in collaboration at the University of Málaga and the University of Glasgow; and the *Corpus of Scottish Correspondence* compiled at the University of Helsinki.

of the University of Glasgow) and transcriptions to be viewed side by side, as well the ability to do simple corpus searches.

While both of these explicitly linguistically oriented projects—as well as most of the other digital editions mentioned below—also include a user interface for the edition, the analytical tools provided by it are not very sophisticated from a corpus-linguistic point of view. Fortunately, in addition to linguistically oriented digital editing projects, the *Textométrie* project—led by Serge Heiden at the ENS de Lyon—shares the aim of facilitating the linguistic analysis of historical texts and furthers it by developing the *TXM* software suite. It is a free, open-source corpus analysis environment and graphical client that is based on the same technologies (Unicode and TEI XML) that are used by the present edition (see section 5.7) and employs established corpus search and statistical analysis engines (CQP and R), making it the prime candidate for integration with an online interface for the present edition.<sup>82</sup>

Grund (2006: 118-9) mentions the Society for Early English and Norse Electronic Texts project<sup>83</sup>, *The Canterbury Tales Project*<sup>84</sup>, *The Electronic Beowulf* project<sup>85</sup>, along with some smaller scale projects like the edition of Henry Machyn's diary<sup>86</sup> as indications of the tremendous progress, from the point of view of a historical corpus linguist, that digital editing has made over the last decade and a half. All of these editions aim at making available all manuscript versions of a text instead of just one canonical text, making them usable as linguistic evidence.<sup>87</sup> Also Lass (2004b) argues that it is possible to “produce a corpus whose basis is edition-free, but still highly informative, relatively unbiased but still with a rich analytical and heuristic apparatus, and with maximal freedom for manipulation by users” (39–40). He then mentions two corpora that meet his conditions, both being prepared at the University of Edinburgh: the Linguistic Atlas of Older Scots (LAOS-1)<sup>88</sup> by Keith Williamson, and the LAEME<sup>89</sup> by Margaret Laing and himself. To these two might be added the *MEG-C*, whose compilers aim “to record what is visible in the manuscript, rather than giving editorial interpretations” (Stenroos and Mäkinen 2009: 14), reproducing the text “at what might be called a rich diplomatic level” (7).

As was mentioned above, the *EWD* also explicitly sets out to answer “the recent call for more linguistically-oriented editions” (Kytö, Walker and Grund 2007: 66), combining features of a linguistic corpus to those of a digital edition while still identifying itself explicitly as an edition. In describing the *EWD*, the editors have introduced the concept of a *linguistic edition* and defined it as an edition where “the language of the original manuscript text is not normalised, modernised, or otherwise emended”, but “the manuscript is reproduced as closely as possible in

<sup>82</sup> Since the development of a user interface is beyond the scope of this thesis, it will be undertaken as a separate postdoctoral project.

<sup>83</sup> <<http://www.iath.virginia.edu/seenet/>>

<sup>84</sup> <<http://www.canterburytalesproject.org/>>

<sup>85</sup> <<http://ebeowulf.uky.edu/>>

<sup>86</sup> <<http://quod.lib.umich.edu/m/machyn/>>

<sup>87</sup> In fact, Curzan and Palmer (2006) describe the digital edition of Henry Machyn's Diary as “a corpus of one text” (27).

<sup>88</sup> <<http://www.lel.ed.ac.uk/ihd/laos1/laos1.html>>

<sup>89</sup> <<http://www.lel.ed.ac.uk/ihd/laeme1/laeme1.html>>

transcription” (Kytö, Grund and Walker 2007: section 3).<sup>90</sup> While the basic editorial and transcription principles of the edition are very similar to the current edition, the technical solutions unfortunately follow rather traditional corpus linguistic conventions, being modelled after the *HC*, published already in 1991.<sup>91</sup> The principal shortcoming of the *EWD* from the point of view of this thesis is that it does not properly make use of the dynamic digital environment by separating data from its presentation (see chapter 5) but bases its multifunctionality on two separate, static versions—one for reading and another for searching—generated from an initial raw transcription before publication (Kytö, Walker and Grund 2007: 75). This static and parallel format means that it is both more difficult to maintain and extend and less flexible to use than the kind of multi-stage edition described below in chapter 5 and section 5.4.<sup>92</sup>

Despite its literary origins, the *Women Writers Project (WWP)*,<sup>93</sup> a multi-genre textbase of English writing by women from 1330 to 1830, has been designed equally as “a resource for social and cultural historians, historical linguists, and other disciplines” (Renear 2001: 36). To this end, it was—like the present edition—designed as a non-critical editing project producing diplomatic editions not merely out of convenience, but because the editors “wanted to support textual scholarship, and not preempt it; that is, provide relatively neutral resources that could be used by other scholars, even scholars who had deep disagreements about the text” (36). Another large-scale diplomatic edition that could be characterized as a textbase is the digital edition of Wittgenstein’s *Nachlaß*<sup>94</sup>, produced by the Wittgenstein Archives Bergen (WAB) and described by Ore (2004). Like the *WWP*, it is not a critical edition in the traditional sense, but rather a heavily encoded digital diplomatic edition which also integrates a normalized reading version, representing “an enormous amount of intellectual scholarly effort” (36-7) that can also be used as the basis of further research and editorial work.

Smaller digital editions—either completed or in preparation—that fulfil at least some of the requirements mentioned above include for example Grund’s so far unpublished edition of an alchemical multi-version text known as *The Mirror of Lights*, which was conceived as a direct response to the call by Lass (2004b) for using early English manuscripts as the basis of linguistic research (Grund 2006: 119). Like the present edition, it is intended to “be useful not only for linguistic studies of various kinds but also for studies of the development and adaptation of alchemical theories and practices” (119). In a very similar vein, the edition of the medieval poem *La Belle dame qui eut mercy* described by Grenier-Winther (2004) is also intended to provide the reader with full access to the text, “from facsimile view of the original manuscript folios to diplomatic and critical transcriptions of each

<sup>90</sup> While this definition does not cover all of the requirements posed above for a corpus linguistic edition, the principles underlying the present edition and the linguistic edition are essentially the same.

<sup>91</sup> Incidentally, the *HC* has recently (2011) received a technological upgrade into a TEI XML encoded version (*HC TEI XML* 2011).

<sup>92</sup> Thus, although Kytö, Walker and Grund (2007: 83) claim that the *EWD* will “surpass printed editions” by being “larger, more accessible, more flexible, and more faithful to the original manuscript”, its persistent adherence to somewhat outmoded conventions of the corpus linguistic community prevent it from fully utilising the potential of current digital technologies

<sup>93</sup> <<http://www.wwp.brown.edu/>>

<sup>94</sup> <<http://www.wittgensteinsource.org/>>

manuscript witness to the text”, along with “the ability to conduct statistical and summary searches on the components of the poem” (193). Although Johansson (2004) does not use the term *linguistic edition*, he mentions linguists as one of the main target audiences of the *The manuscript texts from Vadstena monastery in digitized form* project, and the editorial principles and practices he describes are very similar to those employed here. Another editorial project producing digital editions that provide most of the functionality required for corpus linguistic study is the *Digital Archive of Letters in Flanders (DALF)*<sup>95</sup> project at the Centrum voor Teksteditie en Bronnenstudie at Ghent. Unfortunately, none of these editions—most of them begun a decade ago—are as of yet available for scholars.<sup>96</sup>

Of digital editions actually available for the general scholarly public, perhaps the best from a linguistic point of view is the *Letters of Clemency from the Chancery of Brittany* by Nicole Dufournaud.<sup>97</sup> Despite its age and somewhat dated technology (the edition was finished in 2003), the edition not only provides the online user with facsimile images, diplomatic transcripts and indexes, but also allows users to download the entire edition for use with external analysis software. Despite its technically very simple interface, this feature is enough to make it valuable for corpus linguists and demonstrates that meeting the needs of corpus linguists is as much a question of policy as it is of technology. At the opposite end of the spectrum, Carlquist (2004) has envisioned a new kind of philological edition, which would contain not only high-quality colour facsimiles of the original manuscript and a detailed transcription of its textual and paratextual content, but also a vast array of links both within the edition and to all the passages in other texts that are referred to, as well as contextual information about people, places and cultural phenomena, as well as “grammatical, lexical and phonological explanations of certain passages” (115-6). Unfortunately, it has one more feature that is typical to editions of such scope and complexity: it seems to have never been completed.<sup>98</sup> Cautioned by the paucity of digital linguistic editions that are actually available, one of the principal aims of this thesis is to outline a model for editions that are not only optimally suited for the purposes of historical corpus linguistics, but can also be made available to the scholarly community in a reasonable time by a small project or even an individual scholar.

## 4.4 Conclusion

At present, XML offers a way forward, enabling more complex document models to be delivered over the Web. In the meantime, I would like to reiterate my plea for more basic research and development on what an electronic edition should look like and on how it can

<sup>95</sup> <<http://ctb.kantl.be/project/dalf/>>

<sup>96</sup> The DALF project has produced a set of guidelines and software tools, as well as several pilot editions of correspondence materials, but even they are unfortunately not yet available online.

<sup>97</sup> The edition is freely available online at <<http://nicole.dufournaud.net/remission/>>.

<sup>98</sup> While Carlquist is undoubtedly correct in saying that “such an edition cannot possibly be the work of a single person”, I cannot help but wonder whether such an ambitious undertaking would be feasible even for a larger research team, without being broken down into several consecutive stages.

best maintain and enhance current standards of editorial scholarship.  
(Hockey 2004: 374)

In answer to the plea by Hockey and others, the present edition and the description of its editorial principles and encoding and annotation practices in chapters 10 and 11 attempt to define a model for representing the linguistic, textual and paratextual features of the original documents in a form which can be analysed using corpus linguistic methodology. The design of the present edition has been guided primarily by the requirements of historical variationist corpus linguistics. It is based on a thoroughly New Philological view of textuality, in which the primary object of interest is not the abstract *work* or even the disembodied *text*, but rather the *document*, which is seen as a physical vehicle of communication and the material embodiment of a linguistic and communicative act. In addition to its textual content, meaning is also seen to be embedded in its material paratext, the visual and material choices made in realising the text as a physical document. The term *digital edition* is here not taken to refer simply to the publication medium of an edition. An edition that is published in a digital format but simply replicates the essential functionality of a printed edition is not here considered to be a digital edition but merely an electronically printed one. In order to be considered a digital edition, an edition should not only be published in a digital form but to expand upon the capabilities of the printed edition in one or more of the following ways:

- including more material, either in terms of edited text or annotation, than could be presented in print,
- allowing the presentation of all variant versions of a text in parallel so that all readings of a passage can be conveniently be compared with one another,
- presenting the transcriptions and all annotation, i.e. both data and metadata, in a format that allows them to be searched, counted and related to each other using computational methods,
- allowing the user to access all the information that is available to the editor in a format that allows it to be manipulated by the computer for a variety of purposes (generation of wordlists, reorganization of the text, extraction of subsets of the data, etc.), or
- allowing the user to extend the data and metadata of the edition by adding—preferably through linking and aggregation—either new data in the form of new related documents or new metadata in the form of new kinds of annotation, and to share this new data with others.

Digital editions are very similar to textual corpora in practical terms, differing mainly on the conceptual level. Despite this similarity, which makes it quite easy to accommodate the requirements of corpus linguistics in the preparation of a digital edition, very few corpora have so far been compiled from original documents, most being based on traditional printed editions, mostly for convenience. This thesis, however, argues that considering our current understanding of the communicative significance of the original document, we should no longer settle for compiling historical corpora from printed editions that are not designed for linguistic use, but rather from original documents by editing them in a manner that takes into account the requirements of the discipline. By developing efficient and

purpose-designed encoding schemes, tools and working practices, we can minimize the difference in the amount of work between digitising an existing printed edition and editing an original document, so that the benefits of the much richer data contained in a corpus linguistic edition like the present one will outweigh the remaining difference.

The compilation of corpora from original documents is also supported by the fact that for Middle English nonliterary texts—which linguists are mostly interested in—there is a very limited number of pre-existing editions that have not already been included in corpora, and in the present age, there really should be no excuse for not publishing a new edition in digital form (even if it is also published in printed form). Equally, there should be no good reason for editors not to adopt editorial practices that take into account also linguistic requirements, since fulfilling the requirements outlined above—mainly by *not* doing certain things and including also ‘intermediary’ products of the editorial process in the edition—does not significantly increase the work required by the edition. For this reason historical linguists should also be more vocal in lobbying for the editorial and digital humanities communities to take their needs into account in designing new textual resources. By producing corpora out of original documents we would not be limited to republishing work that has been done by earlier editors but bringing into existence genuinely new research resources—useful not only for linguists but also historians and other scholars—that could be used as the starting point for further work.

A digital edition intended to serve as data for historical corpus linguistic analysis—or a *corpus linguistic edition* should not only present a machine-readable documentary transcription of the textual content of one or more individual historical documents, but should also encode the material and logical structure of the document in a machine-readable way to allow linguistic features to be analysed in relation to their documentary and textual contexts. Furthermore, a corpus linguistic edition should indicate explicitly the source of each textual feature, i.e. whether a passage has been written by the original scribe, a corrector, a later annotator, etc., to allow the separation of different scribal idiolects in the analysis. On the level of usability, a corpus linguistic edition should provide flexible access to the data, allowing the user access to the text and all levels of annotation, which should be encoded using a recognized open standard language like XML to allow their analysis using a variety of tools. Furthermore, as an extremely important feature that is usually not implemented in digital editions and very rarely even in linguistic corpora,<sup>99</sup> the edition should allow not only the exportation of material from the edition for analysis, but also the addition of new analytical metadata into the edition based on the results of this external analysis, i.e. the ability to integrate the results of new research into the research resource itself. And finally, as should be obvious to anyone designing a scholarly research resource, both the sources used for the edition, i.e. the documents edited, and the principles and methods for modelling them in the edition should be documented and described in detail.

<sup>99</sup> A notable exception to this is the *British National Corpus (BNC)*, whose online interface, the *BNCWeb* (based on the *Corpus Query Processor* of the *IMS Open Corpus Workbench*) which allows not only the exportation of search and concordance results, but also the uploading of external files containing e.g. manually analysed and filtered search results to serve as the basis of further analysis, based on textual location identifiers in the original XML data files.

For the production of such editions to be possible not only for large-scale institutional projects but also for individual scholars or small projects, they should not be produced as monolithic and closed products by a single entity, but rather broken up into separate but interlinked components that can be produced in a distributed fashion. In order to promote this, the following chapter will describe an ontological model for a digital edition that separates the edition as structured collection of data about the original textual object from the edition as a research tool, represented by its user interface, and argues for the separate development of these two aspects of the edition. Similarly, it will also argue for the delimitation of the editorial task to the documentation and description of the textual document, leaving its higher-order analysis to the linguistic, literary or historical scholar. This kind of a modular approach, supported by well-defined standards and documented practices, should allow for the agile and distributed development of the individual components required by a digital *research environment* and for the cumulative production of analytical data related to the textual objects.





## Chapter 5

# Ontology of the digital edition

Although an electronic edition has often been seen as merely a form of presentation which does not pose any new theoretical issues, Vanhoutte (2006) has pointed out, that “the practice of creating an edition with the use of text encoding calls for explicit ontologies and theories of the text that do generate new sets of theoretical issues” (163). While the basic methods used by editors to produce digital editions are similar to those used by print editors, the end result itself is very different from the static and ‘flat’ typescripts and word processor files of a print edition, simply because while the final typescript of a traditional edition “had to be intelligible to human readers; the electronic edition, on the contrary, must be intelligible to a computer” (Baker 1998: 273-4).

This change in what Bolter (1991) has called the *hard structures* of the edition—the “tangible qualities of the materials of writing” (41) in which the text is stored and through which it is experienced, can justifiably be called “the greatest single technological change in the history of writing” (Bolter 1991: 42). In print, the physical storage medium of the edition is also its user interface, and their relationship is fixed, both physically and ontologically. In a digital edition, its physical storage medium of the edition—computer memory or a digital storage device of some kind—is decoupled from its user interface, most commonly a computer monitor or some other digital display device. Furthermore, unlike the printed page, the storage form of a digital edition is not directly accessible or perceptible by the human user, but always requires an intermediary filtering operation performed by computer hardware and software, effectively abstracting the text from its reader (and writer) and allowing a multiplicity of relationships between the abstracted storage form of the text and its presentation:

The revolution of the electronic text will also be a revolution in reading. To read on a screen is not to read in a codex. The electronic representation of texts completely changes the text’s status; for the materiality of the book, it substitutes the immateriality of texts without a unique location; against the relations of contiguity established in the print objects, it opposes the free composition of infinitely manipulable fragments; in place of the immediate apprehension of the whole work, made visible by the object that embodies it, it introduces

a lengthy navigation in textual archipelagos that have neither shores  
nor borders. (Chartier 1995: 18)

Although the remediation of the text from a codex to digital format has been seen as doing “violence to the texts by separating them from the original physical forms in which they appeared and which helped to constitute their historical significance” (Chartier 1995: 22), this criticism does not actually apply to editions of manuscript materials, as the same violence is also done—to an even greater extent—by a printed edition. As Chartier (1995) also points out, this kind criticism often makes the mistake of assuming that the digital edition is intended as a replacement of the original document:

A digital representation of an artefact is a representation of certain relevant characteristics of the artefact. It is not the original and complete artefact, nor even a metonymy or *simulacrum* of the complete artefact. (Arnold 2008: 158)

Dahlström (2009: 35) has summarized the “two recurring fallacies in digitization debates and media theory” that underlie this mistake, originally defined by Willard McCarty on an e-mail discussion list: the “complete encoding fallacy” and the “mimetic fallacy”. The first of these was defined by McCarty as “the idea that it is possible completely to encode a verbal artefact”, and the second as “the idea that a digitized version will be able to replace its non-digital original”. From the point of view of historical linguistics, the untruth of these claims is actually the fact that makes digital editions such powerful tools: digital encoding forces the continuous and analogue nature of the original document to be reduced into discrete and digital categories that endows it with superior analytical power with respect to those phenomena that are encoded in it. This means that instead of replacing the original, a successful digital edition ideally surpasses the usefulness of the original in terms of *some specific research questions* while being more or less useless for others. In the corpus linguistic context, digitization and annotation of a document can be seen as a way of *qualitatively* analysing the document in terms of predefined categories and recording the results of this analysis to be used for *quantitative* analyses that would not be possible on the original document.

The first section of this chapter describes the ontology of the digital edition in terms of the separation of data and presentation, while the second one evaluates the implications of this ontology on the status of the edition as a finished *product* versus a ‘work-site’ for an ongoing process. The three following sections proceed to describe the realization of this ontology through the use of *annotation* as a means of modelling the original document, discussing also the ‘economics’ of annotation and the use of separate levels of annotation to represent different ontological levels of the original document. The chapter finishes with a brief introduction to the basic concepts of the XML technology and the TEI Guidelines for Electronic Text Encoding and Interchange that are used for the practical implementation of this scheme.

## 5.1 Separation of data and its presentation

As Johansson (2004: 93) observes, discussions about digital editing often focus on what we see on the computer screen, i.e. the presentation of the edition, instead of the underlying encoding and annotation systems that make it possible to use the edition for something more than simply reading it. Although it is the output of a digital edition that would conventionally be perceived as ‘the edition’, Pierazzo (2011) argues that it is actually the *source file*, stored in computer memory, that “can be considered an edition by itself” (474).<sup>1</sup> In fact, the presentation of a digital edition—such as the ones included in appendices B, C and D of this thesis—does not need to have a permanent existence, but could also be merely the ephemeral result of applying a set of processing instructions to the source file, persisting only for the space of a single viewing of the text. Sperberg-McQueen (2009, originally written in 1997) has conceptualized this multifaceted ontology of digital editions by defining two distinct kinds of knowledge involved in its production: “a knowledge of facts normally transmissible by language (German *Wissen*)” and “the knowledge of how to do something (German *Können*)”, or *knowledge* and *capability* (30-1). In more concrete terms, the knowledge of a digital edition refers to the database of textual and paratextual content with descriptive metadata that encodes various aspects of the original document, while the capability of the edition refers to the edition’s capacity to instruct the computer to do things with that knowledge, manipulating it and presenting it to the reader through a software interface.

This distinction between knowledge and capability also has implications for the status of the digital edition as a bibliographic object. While in the analogue world, the printed book as a physical document embodied and delimited the edition as a textual object, and was “all that libraries and librarians need to know in order to collect and preserve it” (Deegan 2006: 359), the material existence and hence the preservation of a digital edition is more complicated. Kirschenbaum (2002) has argued that although electronic texts do not enjoy a material existence in the sense that you could touch them, this does not make them immaterial in a wider sense. He sees them to exhibit a dimension directly analogous to McGann’s “bibliographic codes” in the form of “the computational variables” of the interface through which they are presented, as well as the data standards and file formats in which they are stored and “which contribute to the textual composition of electronic objects” (43). While both physical and digital documents can both be seen as artefacts in the sense of a man-made object, the difference between printed and digital documents is that while the former is a static object, the latter is a dynamic mechanism “with parts and components, structures and substructures” (42). Thus, while in a traditional printed edition, the document serving as the user interface does not *do* anything to the textual and paratextual content besides visually presenting it in a single way predetermined by the editor, the user interface of a digital edition can *do* a variety of things to the textual and paratextual content with or without user involvement: in addition to visually presenting it in variable ways,

---

<sup>1</sup> This is especially obvious from a corpus linguistic point of view, since the corpus linguist will in many cases never see the edition as a visual ‘reading text’ but rather mine it for information using computational methods and receive for her perusal only the results of a search operation.

it can also filter it, order it and selectively combine it with textual content from other editions.

### 5.1.1 Multiple identity of digital editions

Based on their dynamic nature, Pierazzo (2011) argues that the editions in the digital medium are “substantially different” from printed ones and that the “very definition of ‘diplomatic edition’ needs to be substantially revised” when the edition is published digitally. For this purpose she introduces a new type of editorial object or publication form, the “digital documentary edition” (463), defining it “as the recording of as many features of the original document as are considered meaningful by the editors, displayed in all the ways the editors consider useful for the readers, including all the tools necessary to achieve such a purpose” (475). The conceptual model for this kind of an edition can be seen to consist of four components (Pierazzo 2011: 473-4):<sup>2</sup>

- 1) Source file containing the transcribed and encoded text (and paratext) of the original document(s), i.e. the *data* (including metadata) of the edition.
- 2) Scripts that represent given aspects of the source file in a given way, providing specific *views* of the text.
- 3) One or more outputs, each of which represents one of the views defined by the scripts, possibly including a diplomatic one; this is what would in the traditional parlance be called the *edition*.
- 4) One or more stylesheets that refine the visual display of the edition.

While the output is the component perceived as ‘the edition’ by the user/reader, it is the source file that stores the editor’s knowledge and understanding of the original document (Sperberg-McQueen’s ‘knowledge’), while “it is the scripts that store the knowledge (the scholarship) of how to produce such an edition” (Sperberg-McQueen’s *capability*) (Pierazzo 2011: 473). The solution to the ambiguous identity of the digital edition suggested by Pierazzo is to see the *documentary digital edition* as comprising “all three components of a digital publication—the source, the output and the tools to produce and display it”, all of which “are scholarly products that result from editorial practice” (Pierazzo 2011: 474-5). However, this equation of ‘the edition’ with the complex object containing all of the above components is problematic, because it implies a stable ontological association between the three aspects of the edition (source, tools and output) where there in fact need be—and as I will argue below, should be—none.

Returning to the distinction between knowledge and capability made by Sperberg-McQueen (2009: 30-1), we can see that while the source file, representing the former, defines the identity of the edition as an edition *of a specific textual object*, the scripts used to process the data and the stylesheets used to display it, representing the latter, define the identity of the edition as a specific *kind* of edition. As the result of the interaction of these two, the output, then, represents *a specific kind of edition of a specific textual object*, analogous to a traditional printed edition. Thus the edition in the collective sense used by Pierazzo in fact consists of

<sup>2</sup> From a technical standpoint, the first component is—both in the present edition and Pierazzo’s description—represented by a TEI XML document, the second by XSLT scripts, the third by HTML documents and the fourth by a CSS stylesheet.

two entirely different kinds of components, the *data archive* describing the original textual object, contained in the source file, and a *processing engine* used to generate an editorial output from the data, consisting of not only the scripts and stylesheets but of the application framework that executes them. This means, as Pierazzo (2011: 475) acknowledges, that there is ample room for terminological and conceptual confusion, as ‘the edition’ can quite justifiably be taken to refer either to the source file or data archive, the editorial output, a combination of the data archive and the processing engine, or the entire complex made up of the data archive, the processing engine and all the possible editorial outputs that can it can generate from the data contained in the archive.

While it is possible to define new concepts for each of these components and their combinations, as Pierazzo does in defining the whole complex as a “new type of editorial object, the *documentary digital edition*” (2011: 475), and as I have tried to do above by introducing the terms (if not the concepts) of *data archive* and *processing engine*, it is difficult to overcome the tendency of textual scholars to try and pin the identity of ‘the edition’ on one of these concepts in exclusion of the others. On the other hand, considering the whole complex as ‘the edition’ runs the risk of associating a given data archive exclusively with ‘the’ processing engine ‘of that edition’, essentially locking the data within a specific presentational and analytical framework and not only diminishing its flexibility (see subsection 4.2.1) but also precluding the communal development of new tools for using the data and causing the whole edition to age with the technology of its user interface.<sup>3</sup>

From the point of view of corpus linguistics, which sees the edition primarily as a source of research data, it is most natural to locate the identity of the edition in the data archive or source file, and to see the scripts and other processing and presentation tools as something that are ‘used on’ or ‘applied to’ the edition to produce a specific view or *presentation* of the edition. This formulation is intended to emphasize the independence of the abstract analytical description of the original textual object—the textual and paratextual content and structure of its documentary manifestations and their relationships to each other and to the work—from any specific technical framework used to present or analyse it. This thesis thus agrees with Flanders (2009), who argues “that the kind of edition which best models the text for our scholarly use is something resembling an electronic archive, but one in which analytical and editorial relationships between texts are represented by encoding and by computational relationships” (60), enabling them to be analysed, interrogated and visualized using a variety of different tools.

In order to achieve as complete a separation as possible between the data and its presentation, the present edition takes the following practical steps, recommended by Sperberg-McQueen (2009: 34-7):

- recording the content of the edition in *declarative* terms, using encoding that is optimized for the accurate description of the original document instead of its presentation in a given media;

<sup>3</sup> For example Peter Robinson (2009, originally written in 1999) has, however, argued for the integration of user interfaces to editions and against the creation of mere electronic archives of historical texts in the hopes that people will use them, with the rationale that we must persuade people to use our editions by publishing them as products packaged “in the most attractive form possible” (49–50). This *product-oriented* (see section 5.3) thinking is most likely influenced by the fact that he is producing editions for sale in a commercial context.

- further separating the presentation component into a declarative description of the desired presentation and a separate generalized software environment which provides the actual user interface, guided by the declarative description; and
- using existing publicly documented markup vocabularies (like TEI) for the declarative descriptions of both data and the presentations.

The most important practical requirement that the separation of data and presentation poses for the editor is the requirement to describe the original document in a declarative instead of an imperative fashion in order to keep it independent of its processing. However, as Sperberg-McQueen (2009) points out, data and its eventual presentation are in practice never completely independent of each other, as “the capabilities offered by the presentation layer of an electronic edition depend in part upon the selection of facts represented in the edition” and conversely, “an editor’s selection of information to record in the edition may be influenced by the interface foreseen for the edition” (33).

### 5.1.2 Benefits of separating data from its presentation

The benefits of separating data from its processing and presentation have been recognized widely, and it is a common practice not only in digital editing, but also in information retrieval, web design and development, word processing, desktop publishing, and model-driven software development. While all of the benefits of digital editions listed in subsection 4.2.1 can to some degree be seen as deriving from the separation of data from its presentation,<sup>4</sup> its most obvious benefit is the added flexibility of being able to present the same data in different ways for different purposes. There are, however, a number of less obvious benefits. First of all, it removes the need to compromise between the faithful and detailed description of the original document and the attractive and readable presentation of its contents (Kirschenbaum 2004: 532-3), as the storage and presentation formats of the edition can be separately optimized for their respective purposes, the former for maximal information retention and the latter for maximally effective visualization of that information. This also means that the inability of the editor to fully anticipate the needs of her users is less of a problem. Supposing that the data describing the original document is stored in a well-documented ‘open’ data format, unforeseen research needs can often be met by developing new presentational tools without needing to re-edit the data; all the editor needs to concern herself with is including as much potentially useful data as possible, without necessarily having to predict all of its potential uses.

In addition to the benefits of separating data from presentation, it is also necessary for countering one of the main drawbacks of digital editions, namely the ever-present threat of technical obsolescence. As a number of digital editors and scholars have pointed out (see e.g. Sperberg-McQueen 2009 (orig. 1997), Grenier-

<sup>4</sup> For example the ability to include an unlimited amount of material in the edition is based on the fact that it can be presented selectively, displaying or analysing only the relevant data, while the representation of variance is based on the ability to display several alternative states of the text. The analytical power of the digital edition is similarly based on the ability to selectively ignore non-presentational analytical metadata in the underlying dataset when visualising it but simultaneously use it for operations like searching.

Winther 2004, Jensen 2004, Crane 2006, Deegan 2006 and Lavagnino 2006), the user interface and software tools of a digital edition tend to become obsolete much more quickly than the edition itself, which means that designing the data structures and storage formats of an edition around a specific mode of presentation is generally considered to be a very bad idea. Instead, as Sperberg-McQueen (2009) and others have frequently argued, the best insurance against obsolescence is again the separation of the results of editorial work, i.e. the data, from its presentation by encoding it in a well-documented, software-independent and non-proprietary storage format that is agnostic in terms of the technology used to access and process the data.

This approach corresponds to the paradigm known as *process-oriented programming* in software engineering, and its main benefit is considered to be that it “allows programmers to think about applications as sets of processes acting upon logically shared data structures” (Godse and Godse 2007: 5–18). The crucial point in this kind of a model is the *interface* between the data and the algorithms responsible for its processing, an intermediary data model which defines the structure of the *data* on an abstract, content-independent level. This has two benefits: once we have defined this interface, we can on the one hand format various kinds of data to the specifications of the interface, and on the other, program software that can process any data so formatted without prior knowledge of the actual data. In the case of digital editions, this means that any edition that is formatted according to a given interface specification can be processed by tools employing this interface, and any presentation or analysis tools that are programmed to process data through this interface can access, process and present data in any such edition. As Deegan (2006: 362) argues, establishing standard interfaces would not only allow the interoperation of tools and individual editions, but would allow those editions to also “exchange data at some level with other systems” like “other editions, library catalogs, databases, dictionaries and thesauri, and many other kinds of relevant information sources”, both present and future.

### 5.1.3 Interoperability and standards

The kind of universal interchangeability of data and presentation software envisioned above requires an interface that is not only standardized but also comprehensively documented in enough detail to allow the production of interoperable data and processing tools. In practice this means that such an interface must unambiguously account for all those and only those encodings that are allowed in the editorial data and metadata, and any tools for presentation or analysis that implement the interface must correspondingly specify a method of treatment for each and every encoding documented by the interface. The problem with the kind of standardization required by interoperability is that it always involves a tradeoff in expressiveness, as Bauman (2011) explains:

Said another way, to make her document (more) interoperational, an encoder either needs to know the application semantics of interest ahead of time, or needs to stick rigidly to prescriptions (sometimes, but not always, expressed as schemas) that she believes the target application (which may not be known) will be able to understand. But

rigidly sticking to the encoding that (she thinks that) the target application needs, or ‘encoding to the stylesheet’ robs the encoder of the ability to be expressive in ways that (she believes) might not be correctly handled by the target application. Common sense says this could result in, and my experience says this often does result in, encoder’s deliberately using encodings that are substandard, less faithful to the document, or outright incorrect per the encoding language, in order to achieve the “desired” results. (Bauman 2011)

Complete interoperability for all kinds of digital editions in the sense defined by Bauman—i.e. the ability of an edition to be used by a computer process other than the one(s) it was designed for without any direct human intervention—is not possible for processes of any complexity that need to take into account the semantic content of the annotation.<sup>5</sup> For this reason, the editorial framework presented here follows Bauman in aiming at a more restricted form of interoperability that Bauman (2011) calls *blind interchange* of data. This refers to the ability of the data to be accommodated to an unknown process—or vice versa—with reasonable human effort, based on accompanying metadata describing the encoding format of the data in both human and machine-readable forms. This would allow reasonably expressive interfaces to be developed for editions of reasonably similar materials with reasonably similar aims. In the absence of a suitable ‘standard’ interface to which the editor could relate her edition—which is currently the case for most types of editions—the best solution from the point of view of interchange is to document the encodings of the edition in sufficient detail, essentially creating an *ad hoc* interface for the edition which could in time become a standard for certain type of editions.

For obvious reasons, these nascent standards should be based on an established model, such as the *de facto* standard for encoding digital texts defined by the TEI *Guidelines for Electronic Text Encoding and Interchange* (see subsection 5.7.2). While the guidelines are—being intended for encoding literally *any* kind of text—as such too general to define an interface in the technical sense, they do provide a relatively stable shared reference point for more detailed guidelines and help to eliminate unnecessary differences between related interfaces. One of the aims of this thesis and the present edition is to lay the basis for such a ‘standard’ interface for linguistically oriented editions of historical manuscript and early printed texts by documenting the interface of the present edition in sufficient detail that it can subsequently be expanded and generalized for other text types besides recipes.<sup>6</sup>

## 5.2 Digital *archives* and digital *editions*

This issue of separating data from its presentation can also be approached from the more functional point of view of the production and use of the edition. Whereas

<sup>5</sup> As Bauman (2011) points out, all digital objects are interoperable on some rudimentary technical level (i.e. any Unicode-encoded text file can be opened for viewing in a suitable text editor and any valid XML file can be processed by an XSLT processor).

<sup>6</sup> The interface implemented by the present edition is documented both as a detailed human-readable description in chapter 11 and as an accompanying formal *XML Schema* included on the accompanying CD-ROM (see appendix A).



the previous subsection discussed the structural separation of the edition into data and its processing and presentation. Vanhoutte (1999) has made a similar distinction from a functional point of view, distinguishing between an “archival function” and a “museum function” of editing. The archival function is defined as “the preservation of the literary artefact in its historical form and the historical-critical research of a literary work”, and the museum function as “the presentation by an editor of the physical appearance and/or the contents of the literary artefact in a documentary, aesthetic, sociological, authorial or bibliographical contextualization, intended for a specific public and published in a specific form and lay-out”. According to Vanhoutte, many of the shortcomings of printed scholarly editions result from their trying to fulfil both functions simultaneously, attempting to present both the textual evidence and the scholarly argument of the editor on the same page. Instead, Vanhoutte advocates a practice “in which the construction of a digital archive that contains all data [...] differs from and precedes the generation of the edition”, calling it “the archive/museum model” (2006: 163). In this model, the archival function would be fulfilled by a *digital archive*, “showing a relative objectivity, or a documented subjectivity in its internal organization and encoding”, while the museum function would be fulfilled by an *edition* “displaying the explicit and expressed subjectivity and the formal orientation of the editor” (Vanhoutte 1999).

Also Lavagnino (2009) has made a similar-sounding distinction between a digital *edition* and a digital *library*. Of these, the former refers to a scholarly edition published in digital form which provides “more than just a good digital reproduction of some piece of text”, surveys “all relevant sources for a particular work” and attempts to “understand the bibliographic situation beyond the obvious (and perhaps misleading) facts of the text” (63). The latter, on the other hand, is simply a mechanical digitization of texts that “does not involve any analysis” and “is devoted to reproducing existing books, but not to any critical or bibliographical analysis” (63). This lack of critical or bibliographical analysis separates Lavagnino’s concept of *digital library* from Vanhoutte’s digital archive, which can never be completely objective and always involves analytical interpretation.<sup>7</sup> In the context of manuscript documents I find this distinction relatively useless, as such a non-analytical digital library of manuscript texts is simply an oxymoron, as will be pointed out in section 5.4.

While the task of ‘proper’ editing—both analogue and digital—has predominantly been the establishment of a critically edited text (i.e. a ‘museum edition’), the preparation of text archives being seen primarily as a preliminary step towards this aim. However, some scholars like Johansson (2004) have recently shifted the focus to the establishment of a text archive of original documentary texts as the primary editorial function, seeing the preparation of a critically edited text merely as an extension or an elaboration of the primary text archive. This shift has been coupled with the acknowledgement that a digitally encoded text is not a mere reproduction but an analytical interpretation of the original document. For example Ore (2004) has modified Vanhoutte’s distinction between archives and editions,

<sup>7</sup> Judging from his examples and the fact that he sees digital libraries as something that “can be created by workers who have no special knowledge of the material, and indeed may not know the language it is written in” (Lavagnino 2009: 64), Lavagnino’s concept of a ‘digital archive’ seems to cover mainly *EEBO*, *ECCO* and similar repositories of digitized images and simple transcriptions.

preferring to call them “archival editions” and “exhibitory editions” to highlight the fact that also ‘archives’ of historical texts—such as the archive of *Henrik Ibsen’s Writings (HIW)*<sup>8</sup>—are always editorial in nature Ore (2004: 43). The same applies to corpus linguistic editions, which are also by definition *archival*, eschewing editorial subjectivity as far as possible.<sup>9</sup>

This kind of ‘staged’ or ‘layered’ development of editions, where each stage is seen as an independent scholarly process, is here considered to be one of the principal benefits of digital editing, as it allows not only an efficient division of labour between multiple independent scholars and projects, but also results in the preservation of the results of each stage as its own editorial object which can be reused for an entirely different purpose.<sup>10</sup> Ore (2004) also provides an illustrative diagram of the possible relationships between different types of editions, which is reproduced here as Figure 5.1.

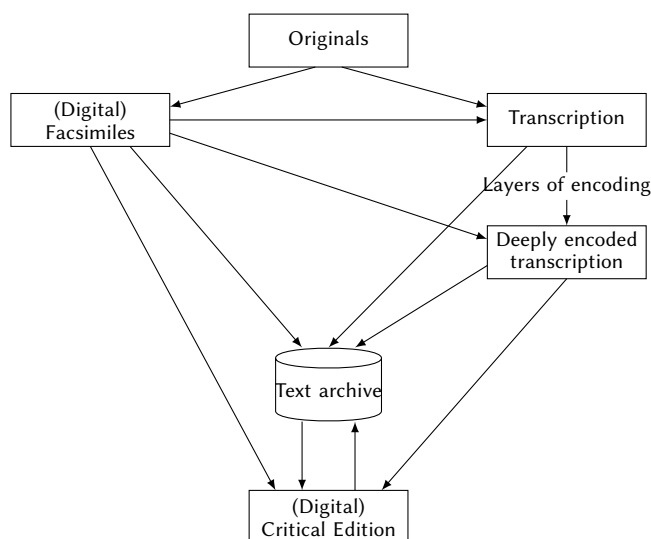


Figure 5.1: Relationships of different edition types, reproduced from Ore (2004: 37).

As Vanhoutte (1999) has observed, the relationship between the archival and museum function of editions is hierarchical in the sense that the archival function is a precondition of the museum function, each digital edition being based on a digital archive of all the relevant documents. Furthermore, the same hierarchi-

<sup>8</sup> <<http://www.ibsen.uio.no/>>

<sup>9</sup> While Ore (2004: 38) still sees the production of a critical edition of all of Ibsen’s writings as the ultimate goal of the *HIW* project, he doubts that they will be the project’s longest lasting contribution to scholarship, becoming outdated long before the archive of transcribed manuscripts.

<sup>10</sup> The *HIW* project is a good example of how editions can be produced on several levels and that it is possible to use the data from one kind of edition for producing another kind (Ore 2004: 37). The project itself has reused data from two earlier projects, a facsimile project of the Ibsen Centre at the University of Oslo and the Ibsen concordance project at the Norwegian Computing Centre for the Humanities Ore (2004: 38), and the archive edition produced by the project can in turn be used not only for the subsequent production of a critical edition but also as a basis of linguistic research and the production of various alternative kinds of editions.

cal relationship should also apply to the editions viewed from the user's point of view: in order to qualify as 'scholarly', every exhibition or museum edition should be supported by access to the archival edition on which it is based and the process of its production documented. In terms or in terms of the structural model presented above, this means that every scholarly editorial product should always be supported and documented by access to both the underlying data archive and the processing engine that was used to produce it.

While some editors, like Flanders (2009), have come to see the production of text archives as a valid form of editing in the sense of "production of knowledge", and as a reason for us to expand "the very idea of the edition" (61), others have criticized this expansion for undermining the editorial role and for replacing the process of "fine discrimination" with one of "maximised accumulation" (Cooper 1998: 92), much in the same way as diplomatic editions have been criticized by critical editors. While the redefinition of the electronic edition as an electronic archive does by necessity represent some degree of "abnegation of authority" (Flanders 2009: 60) on the part of the editor, it is here argued that the whole concept of editorial authority is merely a necessary evil of printed editions, and that most of our scholarly endeavours would in fact be better served by analytically relevant and accurate models of actual historical documents than by editorially authoritative reconstructions of texts that might once have been. Like Flanders (2009: 61), I am inclined to believe that the reluctance to let go of this authority is mainly due to "a concern over our own cultural importance as editors, and over the importance of the work we do", not due to any concern over the quality and usefulness of the scholarly resources we produce.

Accordingly, the present edition and the model of a corpus linguistic edition will be explicitly characterized as an 'archival edition', or "a growing text base" (Vanhoutte 2006: 179) that be used to both mine information about the original documents it models and to generate different kinds of derivative 'museum editions' for both academia and a wider audience, thus providing a tool for diverse research disciplines, including not only linguistics but also history and even critical editing.<sup>11</sup> This means the underlying encoded transcriptions—the result of "major philological effort" (Bøe, Jørgensen and Taugbøl 2004: 70)—are seen as the principal product and the printed diplomatic and parallel reading editions included in the appendices of this thesis are simply examples of two possible outputs—or 'museum editions' generated from the data archive represented by the annotated transcriptions. This kind of an approach has the benefit of allowing new, improved or merely different editions to be generated from the text archive, making use of not only new technological solutions but also of previously unrealized potential in the encoding of the text archive. Like Ore (2009), I feel justified in calling these digital resources *editions*, since they are "products where documents and/or text have been processed and transferred into a new form, usually with some added information and in ways that make the documents more available than they were formerly" (117).

<sup>11</sup> In this regard, the present edition finds its closest parallels in archival projects like the *DALF*, (see Vanhoutte 2006), the digital edition of the writings of Henrik Ibsen (described in Bøe, Jørgensen and Taugbøl 2004) and the Bergen Electronic Edition of Wittgenstein's *Nachlass* (Wittgenstein 2000).

### 5.3 From products to worksites

[T]he electronic medium is perhaps even better suited to preserving variants, and even, with intelligent programming, to letting us see them and study them, not as objects to be valued as they are, but as the record of complex trajectories that do finally lead to works rich and coherent enough to sustain and affirm cultural continuity through times of dramatic change. (Searle 2004: 18)

As the above quotation points out, digital editions—including corpus linguistic ones—are not necessarily any more convenient than printed editions for reading texts, but they are significantly more convenient and efficient for *studying* textual objects. The unsuitability of the ‘electronic medium’ for casual reading was an established trope in the 1990s (see e.g. Lavagnino 1995: 113 and Baker 1998: 263–4), and persisted well past the turn of the millennium (see e.g. Gabler 2006: 344 and McLoughlin 2008: 4). Instead of reading, digital texts were seen useful mainly for “automated analysis, searching, comparison and collation” (Baker 1998: 264), or an “aid to scholarly labor” (Lavagnino 1995: 113). However, since the introduction of Amazon’s Kindle electronic reader in 2007 and Apple’s iPad tablet computer in 2010, it has become clear that it is not the digital medium as such that makes books more or less suited for casual or “sustained sequential” reading (Gabler 2006: 344)—today it is actually easier to “loung[e] beneath a tree with an electronic text” (Baker 1998: 263) than it is with a printed book of any significant size. Instead of the actual medium, it is the way these media are habitually *used* by editors that gives rise to the perceived differences between digital and printed editions.

McLoughlin (2008: 4) has noticed that the rhetoric used by the editors reflects a fundamental difference in the purpose for which printed and digital editions are prepared: whereas printed editions present texts for reading and enjoyment, digital editions present texts as functional objects to be used. However, unlike McLoughlin—himself a print-based editor—seems to think, this difference is not technological but cultural. Or, in other words, not a difference between the media but rather between different types of editions: a *scholarly edition* used as a reference work for study purposes whether in printed or digital form, and a *reading edition*, which could just as well be produced in digital as in printed form (Rehbein 2008: 10). However, as for example Johansson (2004: 95) and Rehbein have pointed out, creating a digital edition just for reading purposes following print conventions does not fully exploit the possibilities of the medium.

The real strength of digital texts is that they allow us to do much more to texts than just read them: we can search individual words or passages, build concordances or do various kinds of computer-aided linguistic and stylistic analyses (Reimer 2004: 176). Thus, in addition to portraying textual objects themselves “as not simply mimetic or expressive, but exploratory, dynamic, and self-correcting” (Searle 2004: 18) through the representation of the diachronic variation inherent in their transmission, digital editions also allow the user to actively participate in this dynamism by offering her access to not only the static endpoint (or hypothetical starting point in the case of critical editing), but to the whole history

of the textual object. The shift of editorial paradigm between printed and digital editions is therefore “not from print to digital — it is from static to dynamic, and it is from output-driven to input- and user-driven” (Rehbein 2008: 2), a change from “a working practice that always had in mind the final result, the book as its output, into a way of thinking that is driven by abstract data, meaning and potential usage” (3). These kinds of digital editions offering not only a presentational publication layer but also all of the evidence underlying it constitute “an interactive resource that assists the user in creating virtual research environments” that make possible “a richer variety of analytical perspectives” and enable “a more holistic notion of what is understood by a text as well as which sources can be represented by a modern edition” (Bodard and Garcés 2009: 96).

### 5.3.1 Relationship between editor and user

With the electronic text, matters will never again be the same. The reader can not only subject an electronic text to numerous processes (index it, annotate it, copy it, disassemble it, recompose it, move it), but, better yet, become its coauthor. The distinction that is highly visible in the printed book between writing and reading, between the author and the text and the reader of the book, will disappear in the face of an altogether different reality: one in which the reader becomes an actor of multivocal composition or, the very least, is in a position to create new texts from fragments that have been freely spliced and reassembled. (Chartier 1995: 20)

This potential of digital editions to empower the reader was also emphasized in the 1990s by Peter Robinson, who saw digital editions as being “much less of the authoritarian editor handing down the definitive text, and much more of a partnership between editor and reader” (Robinson 1998: 261). Whereas a traditional critical edition presents the reader “with the editor’s own fixing of the text [...] and a gleaning by the editor of all the materials surveyed in making this edition” and a traditional multi-text diplomatic edition can usually only present the transcriptions of selected key documents, Robinson (1994); Robinson (1998) envisioned a new kind of edition that would also allow not only the presentation of the editorial vision, but would also provide the user with all the materials and tools used by the editor, enabling her to “test and redo every aspect of the scholar’s work” (Robinson 1998: 261).<sup>12</sup> More importantly from the point of view of corpus linguistics, this kind of an edition also allows the user to do something *completely different* without being burdened by editorial decisions incompatible with that aim.

<sup>12</sup> Some scholars, like Suarez (2000), however, staunchly oppose this kind of ‘abdication of responsibility’, arguing in an incredibly patronising tone that if an editor does not ‘establish’ the text for the user “through the long and demanding process of critical discernment, of carefully weighing complex evidence and deciding in the midst of difficulty” he is “not empowering users but obliging them to arrive at judgments they most often will have neither the time nor the expertise to make” (173). This kind of argumentation portrays the users of a scholarly edition not as intelligent scholars with their own specific research aims and the expertise needed to carry them out, but rather schoolchildren who need to be told not only what the text is but what they should pay attention to.

Grenier-Winther (2004: 191-2) has quite usefully conceptualized the division of responsibility between the editor and user by describing the scholarly edition as an intellectual continuum into which the editor and the reader can enter at different points. In the case of a traditional printed edition, the editor would prepare and present a single ‘definitive’ version of the text (maybe with a list of variant readings), which the reader had to accept. This meant that the editor’s control of the continuum extended far, and the reader entered it near its end, the only task being left to her being its literary or semantic analysis. A digital edition on the other hand allows the editor to grant the reader fuller access to the text and all its variant versions, allowing her to enter the continuum at a much earlier stage. In the case of a corpus linguistic edition, the linguist would preferably enter at the point where the editor has located, transcribed and described the original documents and provided them with analytical metadata about their mutual relationships, but has not yet discarded any of the evidence in order to ‘establish’ a text. Any editorial work past this point—while potentially useful for other purposes—is in the best case wasted and in the worst case makes the linguist’s work more difficult or impossible. Thus I argue, following Grenier-Winther (2004: 192), that in the case of digital editions, it becomes the scholarly editor’s primary responsibility—if for no other reason than scientific verifiability—to ensure that however far she extends her own influence, the user is allowed access at as early a point as possible, and in any case long before the editor’s own influence on the edition ceases.

In practice this means that the only thing the editor of a digital scholarly edition should ‘establish’ is the documentary evidence, recorded as comprehensively as practically possible with “a relative objectivity, or a documented subjectivity” (Vanhoutte 1999: 177), after which point the editor should doff the editorial cap and don that of a textual, linguistic or historical scholar—perhaps better informed than most, but with no inherent privileges as to the analytical examination of the text. All interventions on the documentary evidence beyond this point, whether by the original editor or others should be seen as “nothing more than overlays upon the editor’s work” (Baker 1998: 271) which may filter and annotate the documentary evidence, but not alter it. This kind of an approach would provide the users of *scholarly* editions with the credit they deserve, namely that of being scholars, willing to take on their part of the research and to see the editorially established evidence “as a series of complex textual problems to which they will add their own scholia of commentary, exegesis, and (perhaps) speculative emendation” (Edwards 1998a: 97).

### 5.3.2 Digital editions as dynamic work-sites

Due to the active role in which digital scholarly editions cast their users, they should be seen as digital “work-sites” instead of closed products like printed editions:

The work-site is a text-construction site for the editor and expert reader; and it is the site of study of the work (of its finished textual versions and their annotation) for the first-time reader, as well as any position in between. Because the building of such textual and interpretive work-sites will be, if widely adopted, piece by piece, collabora-

tive and ongoing, we are starting to look at a future for humanities, work-oriented research that is, if not scientific exactly, then more *wissenschaftlich*, in the German sense, than literary critics, historians and others are used to. (Eggert 2009: 81)

As has been pointed out by several digital editors (see e.g. Ore 1999, Johansson 2004, Eggert 2009 and Hajo 2010), the principal benefit of this kind of a “work-site edition” in relation to the print edition, fixed by its moment of publication, is its evolving and accumulative nature, which allows it to “grow as understanding of its textual versions, their physical embodiments and their cultural and historical meanings grow” (Eggert 2009: 82). This, however, will require that editors give up their “anxious, even obsessive” control of the edition and “invite users into the fray as commentators and creators of competing editions using the same repurposed materials” (82). This means that such an edition is no longer as much an end in itself, but should rather be designed to provide an optimal starting point for further research—and even further editions:

[T]he larger goal of an electronic scholarly edition should be not only to meet the current needs of the scholarly researcher but also to stimulate and challenge scholars of various kinds, including teachers, students, even poets and specialists in digital media, to use the text in order to make new knowledge - which is to say, to use it in ways none of us has yet fully imagined. (Fraistat and Jones 2006: 118)

Although envisioned by some editors and corpus compilers already at the turn of the 1990s (see e.g. Rissanen 1989: 17 and Leslie 1993: 49), these kinds of open-ended editions have not become the norm either in digital editing or corpus linguistics, although the technical means have been available for more than a decade.<sup>13</sup> An important reason for this could be the kind of fundamental reconceptualising—and some could argue, diminishing—of the editorial role that their production requires: creating solid, carefully documented but extremely conservative textual archives for the use of others as well as oneself can appear a thankless job, as unglorious as it is useful. However, opening up the insides of our editions and allowing others to ‘interfere’ with them can provide significant boons:

One of the benefits of electronic texts is that, once the most basic editing has been carried out, the creation of further, more elaborate editions becomes both simple and economic. The initial editor’s role as the single, authoritative channel is beneficially diminished: there may be many such channels, and the end-product of none of them need be monumental. I look forward to an era of fundamental editions which present readers with exactly what is there in the source, leaving the power to alter the text in the hands of its ultimate audience. (Leslie 1993: 49)

As (Gabler 2007) points out, the electronic medium allows us to “create an environment of editorial discourses” around the edition “by writing back into the

<sup>13</sup> For example Hunston (2008: 165) clearly sees text corpora as finished products of temporally limited compilation projects as she remarks that once the resources allocated to the compilation of a corpus have been used up, it is impossible to add material to the corpus.

medium the records of what we have analysed and interpreted: discerned, distinguished, and read” (Gabler 2007: 202). This kind of a feedback loop creates a self-reinforcing research environment, where the results of research are placed in direct contact with the object of the research and used to enrich and contextualize it, allowing both the second-order analysis or evaluation of the results in light of the original research materials, and the iterative re-analysis of the research materials themselves in light of the results of all previous rounds of analysis in a kind of hermeneutic circle:

With the appropriate access software, the structuring and organization of the data of the edition should allow computer-based, question-and-answer interaction with the edition. In other words, the electronic medium, instead of being merely an aid in the preparation of a print edition, should become and be recognized and established as the proper site and natural environment for a scholarly edition.

(Gabler 2006: 340)

As one example of such editions Bodard and Garcés (2009) introduce the concept of an Open Source Critical Edition (OSCE), based on a workshop of digital classical scholars. They see OSCEs as “a deeper, richer and potentially different kind of publication from printed editions of texts, or even from digitized and open content online editions”, because “OSCEs are more than merely the final representation of finished work; in their essence they involve the distribution of raw data, of scholarly tradition, of decision-making processes, and of the tools and applications that were used in reaching these conclusions” (84-5, 96). The present edition is based on very similar principles with a similar emphasis on the Open Content model, and could be seen as a close relative of the OSCE model, although tailored to Middle English manuscript materials and to the specific requirements of corpus linguistics.

## 5.4 Modelling the document

We need to reach consensus on a model that will serve current and future scholarly needs and not be overtaken by rapid changes in technology. That model must be flexible enough to address the needs of a variety of users and must also be amenable to many kinds of computer processing. We need to prove that the concept of electronic editions is viable for the long term and formulate a path for arriving at that proof.

(Hockey 2004: 361)

As has been several times mentioned in passing, a successful digital edition of a manuscript document is not merely a *representation* but as an analytical *model* of the original document. This modelling function of the edition is based not only on the transcription of textual content, but on the use of structured markup or annotation to encode various paratextual features of the document and to analytically describe both the original document and its textual content. While transcription



itself was described as a form of analytical modelling in section 2.4, it is in itself essentially linear in nature and can thus encode only a single layer of data (i.e. a single textual stream). As for example Pierazzo (2011) has noted, the use of structured annotation allows not only the encoding of several parallel data streams, but also the modelling of various other aspects of the original document (logical and textual structure, visual layout, physical structure etc.) in a machine-readable form, making them, along with the textual transcription, tractable to corpus linguistic methods.

In this discussion I will take a wider view of annotation than is often the case in corpus linguistic discussions of annotation, taking as a point of departure the appropriate definitions given by the *Oxford English Dictionary* for the verb *annotate* and the noun *annotation*. According to the *OED*, to *annotate* is “to add notes to, furnish with notes” and an *annotation* is “a note added to anything written, by way of explanation or comment”.<sup>14</sup> Along these lines, I will here consider as annotation any kind of explanatory or descriptive metadata added to an electronic transcription of a record of a linguistic event. Essentially, this could be said to include all data about the original document that is not ‘the text itself’, although—as we will see in section 5.6 below—there are levels of annotation that cannot really be separated from the text. This also means that all digital textual research materials by definition contain some annotation, as disconnected and undocumented pieces of text cannot really be called *research material*.

As was explained in section 5.1 above, the crucial difference between digital and printed editions that makes this kind of annotation possible and allows it a vastly enhanced descriptive and analytical power, is their ‘multi-staged’ nature. Whereas in traditional print editing, the editor moves directly from a source feature to its presentational formatting in the finished edition—for example placing an uncertain date in square brackets—in digital editing the encoding of source features involves a two-step process: the editor first encodes the feature based on its meaning (i.e. encodes the uncertain date as such) and then, at the time of presentation, this semantic annotation gets replaced by a presentational representation appropriate for the medium (Rehbein 2008: 8). While this presentational representation might well be the very same square brackets, this intermediary step turns the encoded feature into data, meaning that it can be searched for, counted or processed in different ways. The crucial point with this intermediate stage is—as was pointed out already in chapter 5—that it be strictly declarative instead of imperative, encoding not what the computer should do with the particular textual feature but what that feature is:

Research on the creation and use of markup schemes has illuminated the problems of so-called prescriptive markup, which indicates the functions that are to be carried out on the text. Prescriptive markup restricts the functionality of the electronic text because the text, once marked up in this fashion, can really be used only for the functions prescribed in the markup. (Hockey 2004: 363)

<sup>14</sup> In contrast to this extremely inclusive definition, annotation is often—especially in corpus linguistics—understood more narrowly as referring only to certain types of analytical annotation like part-of-speech tagging or syntactic parsing.

### 5.4.1 Text encoding and annotation as selective modelling

Based on the observation of Sperberg-McQueen (2009)—already discussed in section 2.4—that an edition always records merely “a *selection* from the observable and the recoverable portions of this infinite set of facts” (32) about the original document, Pierazzo (2011: 466) has characterized a transcription or a diplomatic edition as a *model* of the original document. Since scientific models as abstractions must necessarily be simpler than the phenomena they model (Pierazzo and Stokes 2010: 421), the challenge becomes one of selecting such limits to our selection of facts “that allow a model which is adequate to the scholarly purpose for which it has been created” (Pierazzo 2011: 466-7). Any digital encoding of a manuscript text is thus always conditional: phenomena in the original document are abstracted into entities and categorized in various ways with a view to a given purpose. In the case of transcription, graphic marks on the page are abstracted into individual graphemes and can then be categorized into more or less homogenous discrete groups defined according to the purpose of the transcription (Robinson 2009: 44):

It is impossible for a transcription to reproduce the original object; it is always a selection of features from that object: the words but not their size on the page or the depth of the chisel marks, major changes in type style but not variations in the ink’s darkness from page to page or over time. Features that seem essential for a particular transcription can be encoded; what’s impossible is notating everything.  
(Lavagnino 2006: 338)

This means that no transcription of a manuscript text can be definitive or appropriate for all purposes. A manuscript page is “an infinitely complex visual experience”, and a digital transcription of it by default “possesses a clarity and tidiness which hides the complexity of the object of its imitation” (Pidd and Stubbs 1997: 55). Text encoding—like any analytical modelling—is thus an intentionally reductionist activity that “involves defining a closed set of terms by which to describe the text’s structure and behaviour, and applying them consistently” (Flanders 2009: 57). This, however, does not need to imply any essential loss of meaning; on the contrary, it is a strategy of “representing the text according to the qualities we find important to its analysis” (57), or in other words, focusing on the essential. This, as Flanders (2009) observes, has always been the essential task of all editing; the benefit of digital editions over paper ones is simply that they can model a larger number of features in a way useful for a larger variety of purposes.

### 5.4.2 Paratextual transcription

If there is one area in Middle English textual criticism that needs particular work, [...] it is the determination of the meaning and relevance of medieval bibliographic codes. And if there is one area in Middle English editing in need of particular work, in turn, it is the representation of these codes.  
(Machan 1994: 186)

While historical corpus linguists have traditionally focused exclusively on the linguistic content of texts, the field has in recent years shown increasing awareness of the importance of the entire physical document as a communicative vehicle. This has created a demand for corpus linguistic editions that model not only the textual content of historical documents but the material aspects of the document itself. In the field of scholarly editing, the need to record also “the implications of material realization” of Middle English works was recognized by some scholars already in the 1990s, but as Machan (1994: 185) pointed out, this could not be accomplished with traditional printed critical editions. However, many scholars have recently observed that the digital medium has eliminated many of the constraints regarding the encoding and presentation of the documentary or codicological aspects of mediaeval manuscripts.<sup>15</sup> As for example Smith (2004: 312), Fenton and Duggan (2006: 243-4) and Hockey (2004: 362-3) have argued, this is in large degree due to the ability of structural annotation languages like XML to represent various bibliographic and codicological features in machine-readable and manipulable form, making them accessible by the same computational methods that corpus linguists are used to using on the text.

Thus, computers with all their graphical abilities and dynamic presentation of text are actually far better at modelling the functionality of medieval manuscripts than the conventions of modern printed books. While the introduction of digital representation can be seen to take us another step away from the culture of manuscript production, digital editions also allow us “opportunities to present manuscript evidence and various relevant aspects of cultural context more fully and inclusively than the conventions of the printed edition have done” (Reimer 2004: 170-1).<sup>16</sup> Because of the significance of the documentary context for modern historical corpus linguistic methodology, this thesis and the annotation scheme developed in it consider the annotation of material paratext as a parallel operation to the transcription of textual content: as *paratextual transcription*. For this reason, the present edition makes a serious attempt to tackle the challenge posed by Flanders (2006: 140) for digital textual resources, namely to “to capture bibliographic codes and textual materiality in ways that represent them usefully to readers: not simply as visible cues but as data that can give one leverage on the text” (Flanders 2006: 140).

### 5.4.3 Digitization as discrete sampling

In digital editing, the act of transcription—whether of text or paratext—is inextricably bound with the act of encoding it. This “is not necessarily a smooth or painless evolution and [...] comports all sorts of theoretical and practical con-

<sup>15</sup> See for example Cerquiglini 1999, Hockey 2004, Reimer 2004, Smith 2004, Fenton and Duggan 2006, Sutherland 2009, and Pierazzo and Stokes 2010. Sutherland (2009: 20) goes as far as to claim that in fact “the only aspect of the book-bound text that the computer appears to simulate with any high degree of success is the visual”, including many of the features McGann has called *bibliographic codes*: “page size, fount, the placing of the type block on the page, leading, gutters, etc.” but excluding things like surface texture, volume, weight and physical manipulability.

<sup>16</sup> Unfortunately this potential is not always recognized by corpus linguists. For example Claridge (2008: 251) seems to approach the problems posed by historical documents for corpus compilation—such as document damage, unclear writing, additions and deletions and foreign language passages—with the assumption that corpus texts consist solely of plain unannotated text.

sequences” (Pierazzo 2011: 468), which have been examined e.g. by Vanhoutte (1999), McLoughlin (2008), Rehbein (2008) and Pierazzo (2011). Since digital systems use discrete numeric values to represent the “unconstrained and continually varying qualities” (Terras 2010: 50) of analogue phenomena like manuscript texts, digitization always involves the sampling of discrete values from a continuous analogue signal. Consequently, digital encoding of text can be seen as a process of “giving abstract data, letters, spaces, punctuation and so on, a meaning, thus, preparing them for further processing” (Rehbein 2008: 7). Because all sign systems—including the alphabet—have a discrete character, the replacement of any graphic mark on the manuscript page is essentially an act of classification or standardization, of selecting “an individual member suitable to represent it from a finite number of members of a paradigm” (Červenka 1995: 59). As has been generally acknowledged,<sup>17</sup> this classificatory nature of encoding means that neither the transcription of a manuscript text nor the descriptive annotation of its material paratext is an objective or mechanical process, but a highly interpretive one. As with scholarly editing in general, the adverse effects of this unavoidable subjectivity can best be mitigated through the scrupulous documentation of the principles used in making judgements about the text (Bøe, Jørgensen and Taugbøl 2004: 61), which is why chapters 10 and 11 of this thesis have been dedicated to this purpose.

The crucial question for the transcription of both text and paratext is one of granularity, or the number of discrete members in the paradigm that is used to represent the analogue reality of the manuscript. In the age of printed editions the size of this paradigm was limited by the limitations of printing technology; certain aspects of handwritten sources, such as superscript letters, could not necessarily be reproduced in the edition for reasons of economy, forcing the editor to standardize many features of the manuscript (Kline and Holbrook Perdue 2008: 123). However, modern computer-assisted technology makes it economically viable to produce more accurate transcriptions and obviates the need for such heavy standardization. Theoretically computers have entirely obviated the limits imposed by print technology, allowing even a graphic representation of the text, “in which the limitless repertoire of marks in a manuscript is matched by a limitless repertoire of computer signs” (Robinson and Solopova 2006: 2; Kline and Holbrook Perdue 2008: 159). However, such a model of the text—represented in effect by a series of photographic facsimiles of each individual letter in the manuscript—would have very little analytical value, being of equal complexity with the object it models. For this reason, the digital transcription of a manuscript text “cannot be regarded as an act of substitution”, but as a series of acts of “fundamentally incomplete and fundamentally interpretative” translation from the semiotic system of the primary source to that of the computer (Robinson and Solopova 2006: 2). In terms of the different ‘translation strategies’ available to a digital editor for the transcription of the text, Robinson and Solopova have identified four possible options:

- 1) *graphic* transcriptions focus on the concrete, visual aspect of the text, which means that “every mark in the manuscript, every space, is represented in the

<sup>17</sup> See e.g. Parkes (1969: xxix-xxx), Page (1992: 79), Robinson and Solopova (1993: 19), Williams and Abbott (1999: 85), Bøe, Jørgensen and Taugbøl (2004: 61), Berrie et al. (2006: 270), Robinson (2009: 43-4), Walsh (2010a), and Stokes (2010: 239).

transcription, even to the point of decomposition of letter forms into discrete marks”;

- 2) *graphetic* transcriptions, which can be seen to operate one level of abstraction above the purely graphic, organize and classify the marks on the page into distinct “letter-types” (such as the ‘short r’, ‘long r’, ‘round r’) that share the same overall visual shape;
- 3) *graphemic* transcriptions no longer operate on the level of the purely visual but see the different letter-forms as variant realizations of semantically distinct entities called *graphemes*, which means that the transcription encodes all of the variants of what is considered the same letter using the same value;
- 4) *regularized* transcriptions extend the principles of the graphemic transcription onto the level of whole words and represent all variant spellings of a single word with a standardized form of that word.

The selection between these levels involves not only the restrictions of the encoding technology—significantly relaxed in the digital medium—but also the intended use of the edition, the kind of material to be edited, and the resources available to the editors. The lack of constraints also places additional demands on the editor, as a digital transcription can be “searched, analyzed, and edited in ways not possible with a printed transcription”, forcing the editor to cater to a wider range of possible uses (Robinson 1998: 253). While there are arguments for the desirability of minimal granularity for certain kinds of linguistic and palaeographic research, the level of transcription recommended here for corpus linguistic editions and adopted for the present edition is the *graphemic* level, largely for the same reasons as those presented by Robinson and Solopova (2006: 3).<sup>18</sup> The main argument for choosing this level of transcription is that the graphetic approach—which would be theoretically preferable—was found to be impossible to apply with any degree of consistency, as Robinson (2006), Robinson and Solopova (2006) and Rogos (2010) had already discovered in the context of the *Canterbury Tales Project*.<sup>19</sup>

Apart from any considerations of its production, these kinds of “facsimile transcriptions”, have also been criticized for being “too close to the source for easy reading and citation, but too far removed for paleographical studies” (Haugen 2004: 78). In the context of corpus linguistic editing, an additional problem is

<sup>18</sup> Approaching the terminology from the palaeographic point of view of describing scripts and individual hands, Stokes (2011a,b,c) has seen the term *grapheme* as a problematic one, since graphemes “refer only to the abstract and do not in themselves have a physical manifestation” (Stokes 2011c), preferring instead the term *character* for the aggregate group of different letterforms that have the same semantic function. Since the present discussion and the act of grouping individual *allographs* or ‘letter-forms’ into semantically distinct categories for encoding is specifically a process of abstraction, the term *grapheme* is preferred here. The lay term *letterform* is here used as a synonym for the more technical *allograph*, which is used in the same way as in Stokes (2011b) to refer to a culturally ‘accepted’ and established way of writing a certain grapheme in a certain script.

<sup>19</sup> They had originally planned a partially graphetic transcription approach (distinguishing a limited selection of variant letterforms) for the project, but eventually found it to be unrealisable for several reasons: 1) the effort required to distinguish between variant letterforms in transcription caused a significant decrease in the overall accuracy of the transcription, requiring more proofreading; 2) the number of distinct letterforms they could discern increased the closer they looked at the manuscripts, which decreased the confidence with which they could consistently identify these forms across manuscripts; 3) the assumption about a hierarchical relationship between graphetes and graphemes was disproved by cases where an identical graphete was used for different graphemes (identical forms of long ⟨s⟩ and ⟨f⟩ or ⟨c⟩ and ⟨t⟩), which would preclude the automatic conversion from a graphetic to a graphemic transcription.

constituted by the requirement that transcriptions of very different kinds of documents written in different kinds of scripts in different historical time periods would need to be comparable in terms of their transcription practices. This diversity means that it would have been impossible to establish correspondences between the individual graphemes of different scripts from different periods, and thus each script would need its own paradigm of individual and distinct letterforms, making comparison across scripts and time periods impossible in any case.<sup>20</sup>

## 5.5 Economics of annotation

As anyone involved in corpus compilation or the production of other digital text resources knows, all additional information included in the resource always comes with the price of additional work required for its input. Transcribing a manuscript electronically can thus actually be slower than using traditional pen-and-paper methods, “since the electronic text with its potential for markup is capable of conveying far more information than printed editions ever attempted to convey” (Fenton and Duggan 2006: 242-3). While the extent of data contained in a traditional printed edition is immediately apparent upon looking at the page—and possibly turning to the back of the book for indices and critical apparatus—a digital edition can contain much more data and metadata than is apparent from its user interface. As was already observed, digital editions can contain virtually endless amounts of data and metadata. The added functionality and analytical power provided by this additional metadata, however, does not come free, but involves an additional “layer of editing that is both time-consuming and demanding” (McLoughlin 2008: 9). While the extent of a printed edition is in the end constrained by the economic and technical considerations of its publisher, digital editions are rarely published by a commercial publisher, and even when they are, increase in the material included on a CD- or DVD-ROM does not increase the cost or difficulty of pressing, packaging and distributing it. The combination of technological possibilities and the lack of publisher control means that the annotation added to a digital edition often tends to “expand almost indefinitely” (McLoughlin 2008: 7), resulting in delayed and in the worst case, unfinished editions.

### 5.5.1 Conflicting desiderata

Like other factors of an edition, whether printed or digital, the annotation of metadata in a digital text resource like a digital edition or a text corpus is governed by several factors relating to its intended use, source material and the conditions in which it is being produced. These factors include, among others:

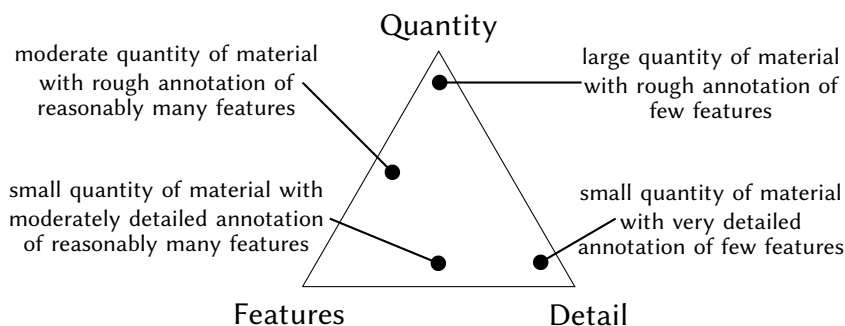
- How much data, i.e. transcribed text, is there?

<sup>20</sup> Considering the diachronic coverage envisioned for the guidelines developed in this thesis, the graphemic level of transcription described in subsection 10.2.2 was found to be a reasonable compromise between accurate representation of the original document and analytical usability. In addition to the transcription of text, an analogous level of granularity is also applied to the annotation of material paratext, annotating features not as specific graphic realization but as (potentially) semantically distinct abstractions.

- How complicated, irregular and idiosyncratic is the original document, i.e. how much metadata is required to accurately describe it?
- Is the edition designed to answer a single well-defined research question or to allow the exploration of a wide range of unforeseen questions?
- What is the institutional context of the annotation process? Is the edition intended for publication or the internal use of a research project? Is it being produced in the context of a legacy project and what kinds of existing practices and conventions are there?

Since the resources of most—if not all—editing, corpus compilation or annotation projects are limited, there is always a trade-off between the depth and breadth of annotation and the amount of data that can be expressed as the following two relationships: 1) the more material there is, the fewer features can be annotated, and 2) the more features are annotated, the less detailed can the annotation be. These two inverse relationships can be visualized as a ternary plot with the three conflicting desiderata of annotation—the *quantity* of data, the number of *features* annotated, and the *detail* to which these features are annotated—located at the points of the triangle. Given fixed resources for the annotation of the edition, the principal decision for the editor thus becomes one of assigning a relative weight for these three factors, depending on the intended audience and the envisioned use of the edition. As indicated by *Figure 5.2*, this decision is always a three-way compromise, more of one thing meaning less of the others.

However, as Hunston (2008: 165) has observed, not all forms of annotation are equal; some forms of annotation can be conveniently added as a part of the transcription process with very little overhead (e.g. representation of special characters and typography), while others require detailed manual analysis of the text and involve a significant amount of work and expertise (e.g. normalization of spelling and part-of-speech annotation for Middle English texts). The amount of work required by certain type of annotation is also dependent on the type of material being annotated, as descriptive annotation of highly standardized printed texts can be automated to a much higher degree than that of irregular and idiosyncratic manuscript texts.



**Figure 5.2:** *An illustration of the relative economics of annotation.*

### 5.5.2 How much to annotate?

In any event there is everything to be said for presenting as much evidence as possible. An approach that selects only those features that the editor thinks significant tends to offer a ‘foreclosed’ account, which cannot be checked without duplicating a great deal of the ground-work already presumably undertaken by the editor.

(Lucas 1998: 178)

In the age of printed editions, scholars transcribed their primary sources for representation in a specific conventional printed form, whose limitations to a great extent defined what was transcribed and how, as there was “no point in transcribing what the printer cannot print” (Robinson and Solopova 2006: 2). Since we have only begun to test the limits of the digital presentation medium, the medium cannot provide a pragmatic answer to the question “where to stop”, as it does in the case of printed editions. Although most editors have perceived this as a challenge, Pierazzo (2011: 467) points out that it is in fact an opportunity for developing editorial theory, as it forces us to find our answer to the question on scholarly and theoretical grounds instead of pragmatic ones. Since a digital edition is not tied to a specific form of presentation and can be used for a variety of purposes, the amount and kind of annotation is not limited by what can be simultaneously presented at any one time. As long as the features of the original textual object are annotated in a way representative of *how it appears in the original document* with no regard to *how it is intended to appear in the edition*, the annotations can be used by a wide variety of different presentations to produce functionality that could not even be imagined at the time of their selection and encoding into the edition (Sperberg-McQueen 2009: 32-3).<sup>21</sup>

As Johansson (2008: 48) reminds us, corpora often end up being used for a much wider variety of purposes than envisioned by their compilers, and a corpus linguistic edition is thus likely to be put to a number of uses that the editor has not anticipated. For this reason, many digital editors have admitted feeling a pressure towards all-inclusiveness, both in terms of data and annotation (see e.g. Leslie 1993, McLoughlin 2008 and Lavagnino 2009). While some editors, like Lavagnino, have seen this as a problem leading to excessively ambitious digital editing projects, others have taken a more positive view:

No one can successfully anticipate the discoveries of succeeding generations, and it is therefore only sensible to be as inclusive as the medium permits. The restrictions on what can be published in traditional form are clear and arise rapidly, because of the economics of print and paper. But the electronic medium offers the opportunity to push back these boundaries. Exploitation of this opportunity will require considerable imagination on the part of editors and constant vigilance to guard against the habit of silent, unconscious self-censorship.

(Leslie 1993: 47)

<sup>21</sup> Sperberg-McQueen (2009: 36) mentions the Wittgenstein Archive at the University of Bergen as a good example of this kind of practice, strictly separating the core of the edition, expressed in a purpose-made declarative markup scheme, from its several presentation interfaces.



Regardless of the number of features we choose to annotate, it is important to keep in mind that there are no features “too obvious to merit recording” (Durusau 2006: 304), as the analytical power of digital editions is based largely on their ability to make “both the obvious and not so obvious structures in a text explicit” (304). While the explicit annotation of a text on both documentary and textual levels undeniably requires more effort than the kind of lighter descriptive annotation traditionally associated with linguistic corpora, it is in most cases likely to be worthwhile, since it opens up multiple views into the text, “allowing users with different interests to choose between several ways of reading the text” (Haugen 2004: 89). While every feature that the editor decides *not* to annotate is forever lost to the user, every feature that she *does* decide to annotate is not only available but also increases the probability for the appearance of *emergent data*, i.e. analytical data that is not based directly on anything encoded into the edition but on the interactional patterns of a complex system of data and metadata:

Indeed, to all intents and purposes there is no limit to the amount of information one can add to a text - apart, that is, from the limit of the imagination. One thing is clear: the more we put into a text, the more we can get out of it. If we are lucky, we wind up even finding things we didn't know were there; if we are very lucky, we find things we didn't even know it was possible to look for. (Driscoll 2006: 261)

However, this should not lead us into believing that it is possible to annotate every aspect of a document. While the digital medium does move the limits of annotation quite a bit further from those imposed by print and effectively eliminates the conflict between readability and quantity of annotation, it does not relieve the editor “of the difficulty of deciding what to omit” or “guard him against possible criticism for having omitted what he should have included” (Gaskell 1978: 6). The challenge lies in striking a balance between the extension of usability that is the benefit of added information and the increase in the editorial work required, which is its cost.

### 5.5.3 Criteria for delimiting annotation

In her recent article, “A rationale of digital documentary editions”, Pierazzo (2011) lucidly describes the multi-staged nature of what she calls “digital documentary editions” and the rationale we should follow in limiting our attempted reproduction of the original document. As a basis for the decision about whether to include a certain feature in a digital documentary edition, Pierazzo (2011: 468) presents a list of five criteria: 1) the purpose of the edition (or the needs of the editors); 2) the needs of the others (prospective readers, scholars); 3) the nature of the document; 4) the capabilities of the publishing technology; and 5) the costs of encoding/the amount of time available for the job.

While it is primarily the processing engine of a digital edition that is responsible for tailoring the editorial output of a digital edition to the needs of different kinds of users, the processing engine can only process what has been explicitly encoded into the edition, which means that the anticipated uses of a digital edition also place requirements on its transcription and annotation principles (Pierazzo 2011: 471). Thus, while the aim of these kinds of editions “has often been defined

as ‘to represent the original manuscript as correctly as possible’”, the definition of ‘correct’ is highly dependent on the use to which the edition is put (Pichler 1995b: 690). Thus, the crucial question determining what to encode in a digital edition becomes: “what does the editor want the user to be able to do” (McLoughlin 2008: 12)? On a more general level, the scope of usage planned for a digital text resource also tends to have a bearing on the level and type of annotation included in it. Resources that are produced for a specific purpose—often a PhD thesis or an individual research project—often receive detailed annotation of a few select features, while more general-purpose resources—such as editions or corpora intended for publication—are more likely to opt for a less detailed annotation of as many features as possible.<sup>22</sup> As Pierazzo (2011) points out, accounting for the needs of the audience is always a balancing act for the editor. In developing the annotation practices for the present edition, I share Pierazzo’s view that while digital editors should be prepared to make significant effort to make their editions as widely usable as possible, they should also avoid spending excessive resources on trying to guess the needs of future scholars and providing specific kinds of annotation for them (2011: 471).

In addition to the needs of the editor and the anticipated users, also the nature of the document may affect the encoding of certain features. For example the layout of a draft document written in a piecemeal fashion and heavily emended might be considered to merit a more detailed annotation than that of a very conventionally laid out clean manuscript copy or a printed version of the same work (Pierazzo 2011: 471-2). The influence of the contextual function of paratextual features on their annotation has also been noted by Claridge (2008) in her discussion of descriptively annotating linguistic corpus texts. She argues that it is necessary to annotate all features that can be considered to be “meaningful”, such as italics used for highlighting, but questions whether this applies also to completely conventionalized typographical features like the italicization of proper names and to visual features like “varying font sizes, spaces, indentation, various types of ornamentation” (253). As potential solutions, she outlines two alternatives. The first one is to see the text as not only a textual but also “a visual object”, i.e. to focus on its documentary nature and represent its visual features without regard to their function, and the second is to annotate only those features “which are of some linguistic relevance”, and possibly those that are “highly prominent” or particularly “easy to encode” (254).

The latter approach, while on the surface a reasonable compromise, has the fundamental flaw of being not only extremely subjective and bound to specific research aims, but also of putting the cart before the horse: it is impossible to evaluate the meaning potential of features before we annotate and are able to analyse them in their textual context. Furthermore, if we make this judgement of relevance on the level of individual documents, the approach has the debilitating effect of producing different annotation practices for different kinds of documents. If, on the other hand, we try to define a general annotation practice on the grounds of

<sup>22</sup> In practice, the usage scope of the resource also determines the thoroughness with which the annotation scheme used needs to be designed and documented. Resources intended for the use of a single researcher or a small project allow for more ad-hoc solutions, as only the annotator herself needs to understand them, while projects that produce public research materials need to spend considerable time and effort in designing and documenting their annotation practices.

relevance, we end up annotating features that are linguistically relevant, prominent or easy to encode in one document but far from it in another. For these reasons and despite Claridge's claim that this kind of an approach would "entail an immense wealth of encoding — and has not been carried through for any corpus" (253-4), the present edition adopts the first alternative, annotating the purely visual appearance of the original document regardless of the functional value of its visual features, seeing the question of their significance as a suitable research question for the users of the edition.<sup>23</sup>

While the digital medium places very few restrictions on what kinds of textual and paratextual features can be encoded in the document, there are significantly more restrictions on what can be satisfactorily presented, even with current visualization technology.<sup>24</sup> For situations where the chosen presentation technology is not able to visualize a feature, Pierazzo (2011: 472) sees three options with regard to annotating it: 1) recording the information in the archival version anyway, in expectation of new technology that will allow its presentation; 2) not recording it to avoid wasting resources on a feature that cannot be represented; or 3) selecting a different output technology that can represent the feature. Since the user interface or presentation stage of editions is currently the site of the most rapid technological development, the present edition follows the first one of these options in that no feature of the original document was left unrecorded because it cannot be visually represented using current technology.

While not a scholarly consideration as such, the institutional and financial conditions under which an edition or a corpus is being prepared also affects the amount of annotation that can be included by setting limits on the amount of available resources, which is—unfortunately—often "the fundamental discriminating criterion" (Pierazzo 2011: 469). Although these kinds of practical considerations should ideally not affect scholarly decisions, it would be naive to deny that the decision whether to encode a specific feature or not is always influenced by the cost—in terms of work-hours—of encoding it. In addition to the economic conditions that it provides for the editorial work, the institutional context also influences the annotation decisions of the edition in other ways. Institutions like research centres, academic departments and scholarly projects often have their own historically established research priorities and scholarly cultures that necessarily affect decisions about what is considered worth annotating. Especially editions or corpora that are produced as parts of a larger whole or within an existing project have precedents that affect—or even dictate—the decisions of what to annotate and how. While this can help create continuity and lead to the development of standards, it can also have the negative effect of outdated practices being carried on past their useful lifespan merely on the force of habit.

Finally, in terms of concrete suggestions on what to annotate, Pierazzo (2011: 467-8) presents a list of features of original documents that might be included in

<sup>23</sup> It should be noted, however, that this does not preclude the annotation of an editorial interpretation of their functional roles, merely that these two aspects are annotated independently of each other. This means that the use of rubrication *as a visual feature* is annotated in exactly the same way regardless of whether it has been used to highlight a number within a paragraph of running text or to indicate a heading.

<sup>24</sup> While it is currently possible to visually represent more features of the original document using digital presentation technology than it is using a printed edition, technologies commonly used for presentation of digital editions such as HTML and CSS do have their—often surprising—limitations.

a digital edition:

- **Documentary features:** dimensions, inks, tears, alterations to the integrity of the physical object.
- **Topology:** structure and layout of the document, collocation of writings and other features in the writing surface.
- **Handwriting:** number of hands, letter shapes.
- **Orthography:** spelling, diacritics.
- **Writing features:** which can be split into
  - **Reading facilitators:** capitalization, punctuation, spacing.
  - **Shorthands:** abbreviations, symbols, cyphers.
- **Genesis:** revisions, deletions, additions, functional marks and other evidence about how the content of the document was produced.
- **Textuality:** paragraphs, headings, verses, tables, lists, rubrics and other structural divisions.
- **Semantics:** dates, names of people, of places, keywords.
- **Linguistics:** part of speech, lemmatization, syntax.
- **Decoration and other graphical components:** miniatures, drawings, doodles.
- **Others:** infinite.

The present edition includes most of the items on this list, excluding:

- a) specific letter shapes (see subsections 5.4.1 and 10.2.2),
- b) semantic features (see section 5.5),
- c) lemmatization and syntactic parsing (see section 5.5), and
- d) an infinite variety of other annotation, which is left up to future users and annotators (see section 5.6).

As the somewhat facetious last item in the list of Pierazzo (2011) reproduced above highlights, a digital edition is never complete—even when the entire document or work is edited, there always remains the possibility of adding more metadata. This unlimited capacity for metadata offered by the electronic medium and the active role of the user change the question of what should be annotated into “what should the *editor* annotate”. Acknowledging the perpetually unfinished nature of digital scholarly editions as research resources and seeing it as their principal virtue instead of a shortcoming, this thesis and the accompanying edition adopt a *layered* approach to annotation, which allows not only the infinite accumulation of new annotation based on pre-existing layers of annotation, but also the division of labour between the editor and the future scholarly users of the edition according to the principle of “expertise” proposed by Roueché (2009): “my intervention should be limited to those matters on which I am uniquely expert, and other experts will take the material forward” (168).

## 5.6 Layered annotation

As Pierazzo (2011) points out, not all of the features listed at the end of section 5.5 are ontologically similar, some corresponding to “graphic evidence on the writ-

ing surface (letter shapes, ligatures, graphic components)” and others representing “meta-information, such as dimensions, or qualifications of words in terms of both semantics and grammatical functions” (468). Based on this observation, the annotation system developed in this thesis for the present edition—and for corpus linguistic digital editions in general—divides the different kinds of annotation commonly included in digital editions and corpora into three categories or layers: *documentary*, *descriptive* and *analytical*. These three layers are not only roughly sequential in terms of the work flow of the editorial process, but also in terms of their decreasing ‘necessity’ in terms of the accurate representation of the original textual object.

The idea of the layered nature of annotation is not a new one; several earlier scholars have also categorized the wide variety of annotation found in electronic research materials in different ways. For example Wittern (2006: 291) has characterized the representation of text in digital form as a three-tiered process, although on a slightly different level from the one discussed here.<sup>25</sup> In the context of language corpora, Claridge (2008) presents a very similar view of corpus annotation as the present edition, seeing it as consisting of “the provision of text headers, textual markup for capturing layout and other surface properties, and grammatical notation in the form of tagging and parsing” (252), which on a practical level corresponds exactly with the three levels of documentary, descriptive and analytical annotation.

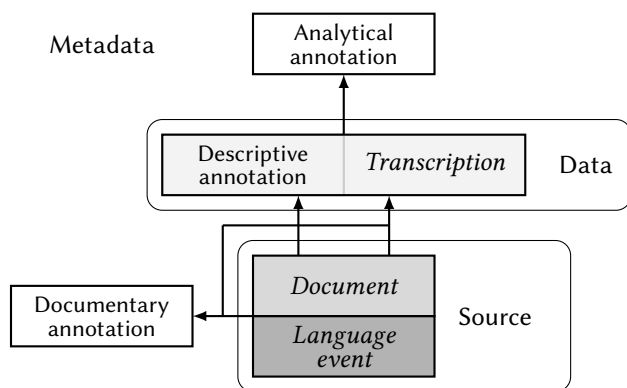
In addition to being associated with different stages in the production of the edition, the different levels of annotation also have differing relationships to the transcription and encoding of ‘the text itself’. The problematic relationship between annotation and the text, namely the question whether annotation (or markup) is a part of the text or something external to it, has been recently outlined by Buzzetti (2009):

If markup is defined, as it has been by the editors of the TEI *Guidelines*, as ‘all the information contained in a computer *file* other than the text itself’, how can it be maintained at the same time that ‘*any* aspect of the text of importance to the researcher’ could ‘be signalled by markup’? For either markup is thought to be information that ‘*is not* part of the text’ and is *different* from the text — and in that case the text is identified with the string of characters representing it — or markup is understood as expressing certain aspects of that information which ‘*is* part of the text, and is *the same as* text’ — and in that case the text is identified with the information content expressed by that string of characters. To overlook that difference is to overlook Hjelmslev’s distinction between the ‘expression’ and the ‘content’ of a

<sup>25</sup> The first layer defined by Wittern is “character encoding”, where each letter of the text is assigned with a taxonomical number value. The second is “text encoding”, where descriptive markup is used to represent the textual structure of the document. The third is “style encoding”, which captures the “shapes and forms that have to be used to re-create the shapes of the letters” from the numbers captured by character encoding. In the present scheme, which separates the encoding of textual and paratextual content, or transcription, from the annotation of analytical structures and categories, Wittern’s first layer belongs to the realm of transcription, while the second one is considered analytical annotation, and the third one—as far as it applies to manuscript texts—is considered a part of descriptive annotation.

text, and to ignore that ‘the representation of any information content is not the information content that is represented by that representation’.  
(Buzzetti 2009: 49)

Buzzetti (2004) himself has come to the conclusion that annotation has a dual character in that it can serve as both “an extension of the writing system itself” or “an extension of the expressive resources of the object-language, i.e. of the very language constituting the text”, and “a metalinguistic description of the structure of the text”, “a form of metalinguistic notation” (178). While agreeing with Buzzetti’s observation, the approach taken here attempts to further clarify the nature of annotation by positing two different kinds of annotation, one of which is ontologically ‘the same’ or a part of the text and encodes features of the original document that are parallel to the text, while the other is *not* part of the text but rather encodes some analytical observations about its structure and meaning. Of the three layers of annotation listed above, it is the middle one, namely descriptive annotation, that is ontologically similar and parallel to the transcription of the text itself, being essentially a transcription of the material paratext of the original document (see page 176), i.e. a part of the *data*—“the raw material deriving from the source itself” (Deegan 2006: 366). The other two layers, on the other hand, can be seen to provide extraneous information that is not derived directly from the original document but can be characterized as *metadata*—“added symbols that describe some features of the data” (366), either in relation to the original document and the cultural and linguistic event it describes (documentary annotation), or to an external analytical framework (analytical annotation). Figure 5.3 illustrates the identities of the different layers of annotation in terms of data and metadata and their relationship to the source, i.e. the original document and the speech event it encodes.



**Figure 5.3:** Conceptual illustration of the different layers of annotation contained in editions and other digital text resources.

### 5.6.1 Documentary annotation

The first stage in the creation of a digital edition involves locating and selecting the material to be edited, which usually consists of a historical document that records a linguistic event that the editor wants to include in the edition. Once the document is located, the usual practice is to find out and record as much contextual information about it—as well as of the linguistic event it represents, i.e. the historical context of its creation, and of the people involved in its production—as can be discovered through archival research and examination of the document itself. This kind of bibliographical and biographical data about the entire textual object that documents the contextual significance of the document, or ‘what the document represents’, is a component of what is here called *documentary annotation*. In addition to information on what the document represents, documentary annotation also includes the documentation of ‘how is the document represented’, i.e. of the principles and practices that have been followed in modelling the document as *data*.<sup>26</sup> As the production of documentary annotation in theory—if not always in practice—precedes the actual editing of the work or document and guides it, it could also be termed *preliminary annotation*.

Since documentary annotation essentially involves documenting one’s sources and methods, it is considered mandatory for good editorial practice, and is usually presented in the form of an introductory essay. As Johansson (2008: 36) observed in his review of the history of corpus linguistics, it has also been considered a mandatory part of linguistic corpora from the very beginning of the discipline:<sup>27</sup>

Background information about the texts and their authors is provided as a rule in the text headers and, ideally, in a comprehensive manual accompanying the corpus which also explains the compilation principles in detail. Background information is vital for both compiler(s) and users in order to ensure and judge respectively the representativeness of the corpus. (Claridge 2008: 252)

While the documentary metadata in editions usually focuses on describing the cultural or literary significance of the work being edited and the codicological features and provenance of the source manuscripts in the form of an essay (in which editorial principles are all too often merely mentioned in passing), linguistic corpora usually focus on outlining the compilation principles of the corpus and providing metadata relevant for the identification and classification of the texts.<sup>28</sup> The present edition tries to combine these approaches, providing the user with both kinds of information, supplemented by a very detailed account of the editorial principles and annotation practices employed in modelling the document.<sup>29</sup>

<sup>26</sup> This kind of documentation of the entire textual object is what is often called ‘metadata’ or ‘descriptive metadata’ (Burnard 2005) in corpus linguistics, in exclusion of information about specific textual units, which in corpus linguistics is often called simply ‘annotation’, and is here divided between descriptive and analytical annotation.

<sup>27</sup> Already the Brown Corpus was accompanied by a manual that explained the sampling and coding of the texts and provided detailed information on the source texts.

<sup>28</sup> These commonly include its title, publication format, register, text type or genre, content (as keywords), style (formal or informal), medium (written or spoken), language use (prose or verse, dialect, foreign languages, etc.), dates of composition and manuscript, and links to records in bibliographic catalogues (Claridge 2008: 252), often provided in the form of formal classificatory parameters.

<sup>29</sup> This information is presented both in a structured XML format in a header included in the source

In addition to its primary purposes of disclosing and describing the sources used by the editor and allowing the user to verify her interpretations, and of supplying the user with information essential for the interpretation of the textual object (Zeller 1995a: 44), documentary annotation also serves other, more practical purposes. First of all, by documenting the identity of the edited textual object, and thus also of the edition itself, it makes the digital representation of the textual object ‘findable’ for researchers (Lehmberg and Wörner 2008: 492).<sup>30</sup> In addition to its immediate use, detailed documentary annotation is also “highly important for the long-term preservation of language resources, as it contains information relevant to the reuse and exchange of the research data, such as legal information, encoding etc.” (Lehmberg and Wörner 2008: 493).<sup>31</sup>

### 5.6.2 Descriptive annotation

After the editor has located and documented the documents to be edited—and preferably also decided upon and at least preliminarily documented her intended treatment of them<sup>32</sup>—the next step is the transcription of the text or the creation of a computer-readable model of the original document, describing it in terms of its textual and paratextual surface features. This should include not just the transcription and encoding of the textual content of the document, but also the descriptive annotation of all the textual and paratextual features of the document that cannot be represented through character encoding, including any special symbols, formatting, highlighting and decoration of the text, the layout of the page, the palaeographical features of the text, as well as any damage suffered by the document, depending on the type of material and the intended use of the edition.

Descriptive annotation is thus parallel to and closely integrated with the transcription of the material, “an extension of the writing system” (Buzzetti 2004: 178) which allows the transcription of not only the graphemes making up the text in the traditional sense, but of also many of its visual and spatial aspects. It is also integrated with transcription in the more practical sense that it is most effectively created during the transcription process, for which reason it could also be characterized as ‘primary annotation’. It also requires access to the original document to an even greater degree than documentary annotation, which is one of the principal features distinguishing it from analytical annotation, described below. Because of its transcriptional nature, the descriptive annotation of manuscript or other complex documents cannot really be separated from the text itself, being instead an

---

file of each text, as described in section 11.1, and in prose form in chapters 9, 10 and 11.

<sup>30</sup> In the context of corpus linguistics, this does not refer merely to the user being able to get their hands on the edition in the first place, but also to the ability of automatically categorising and filtering the text based on its contextual metadata, for which reason it is important that documentary annotation is provided also in a structured format that can be automatically processed by a computer.

<sup>31</sup> For this reason, the present edition also provides a structured formal definition of the edition itself as a bibliographic object (including publication and licensing information), of the historical source document it models and of the encoding and annotation used to model it, using a standard TEI header defined by the TEI Guidelines for Electronic Text Encoding and Interchange.

<sup>32</sup> Although in the spirit of full disclosure, I must point out that the editorial principles and annotation practices of the present edition were not documented until relatively late in the process, as they were still being iteratively developed during the transcription and annotation process—as I suspect is the case in many editorial projects which cannot rely on a previously defined set of principles and practices.



integral part of the transcription as a model of the text and not something added on top of a 'pure text'. Thus, instead of viewing the kind of plain text representation that is commonly found in text corpora and even traditional printed editions as the real text and a descriptively annotated transcription as a 'decorated' version of it, we should instead see the plain text version as a severely impoverished representation of the original document, and the annotated version as a more accurate and complete model of it.

The function of descriptive annotation is identical to that of textual encoding: by classifying the paratextual aspects of the document according to some formal taxonomy and explicitly encoding them in a machine-readable form, it allows them to be not only visually represented but also searched, quantified and analysed by the computer in the same way as the encoded text. Thus, unlike documentary annotation that provides contextual metadata for the interpretation of the textual object, descriptive annotation—together with the textual transcription—constitutes the very data to be interpreted in the light of that metadata.

### 5.6.3 Analytical annotation

Analytical annotation describes the abstract textual object embodied by the document with respect to an analytical framework, encoding information that is not present in the original document. Unlike descriptive annotation, this third and final layer of annotation in the present scheme can in fact be seen as 'something extra', and unlike the other two layers, its production should be properly seen not as a specifically editorial but more generally a scholarly task. In essence, analytical annotation represents the integration of the results of scholarly analysis back into the edition to serve as the basis of further analysis.<sup>33</sup> The common feature of all analytical annotation is that it is based on the underlying editorial layer of textual transcription and descriptive annotation, i.e. the data of the edition itself, describing not what is in the original document, but rather what it means. Because of this, analytical annotation can usually be produced without recourse to the original document, provided that the amount of descriptive data is adequate.<sup>34</sup> Since most kinds of analytic annotation require systematic analysis of the underlying data, it often added to the edition as a separate stage after the initial transcription, constituting what could be called 'secondary annotation'.<sup>35</sup>

The most elementary level of analytical annotation, and generally the only one which is added already during the transcription process, is the division of the text

<sup>33</sup> The kinds of analytical annotation that are likely to be added to an edition naturally depend—to an even greater degree than with descriptive annotation—on the intended use of the edition and the research interests of its users, but can include for example grammatical, syntactic, semantic, discourse functional, metric, or intertextual annotation.

<sup>34</sup> Being essentially the result of research done on the edition, the kinds of analytical annotation that can be produced is limited by the extent of descriptive (and previous analytical) annotation included in the edition.

<sup>35</sup> This distinction between 'primary' and 'secondary' annotation, representing different kinds of editorial interpretation is not unique to the present edition, but was already used as the primary organizing principle of annotation in the digital edition of Wittgenstein's *Nachlaß*, described by Pichler (1995b: 693-5). Although its editors do not conceptually distinguish between descriptive and analytical annotation, they initially included in the edition only those features which required reference to the original documents (description), and left for later annotation which can be inserted with reference only to the transcription (analysis).

into paragraphs, headings, sections and chapters, i.e. the annotation of logical textual structure.<sup>36</sup> This kind of structural annotation—even when relatively detailed and distinguishing not only generic elements like paragraphs and headings but specialized ones like date lines, closings and signatures in letters—is generally considered rather straightforward. Despite this apparent simplicity, Pichler (1995b) has nevertheless argued that although all transcription involves interpretation, the interpretation involved in the recognition of letterforms in strokes on the page or individual words in a stream of these letterforms is “quite different” (Pichler 1995b: 692) from the kind of interpretation involved in the designation of a certain string of words as, say, a title. However, the distinction between transcription and descriptive annotation on the one hand and analytical annotation on the other is here not seen to lie in the act of interpretation itself, but rather on the former encoding observable phenomenon present in the original document (e.g. some words being written with red ink and more formal hand) and the latter encoding an analysis of their textual function.<sup>37</sup>

In the context of corpus linguistics, ‘annotation’ is usually understood more narrowly as the addition of analytical linguistic data. In nearly every corpus that has been annotated with linguistic information, the first level of linguistic annotation is the annotation of the part-of-speech POS of each word, which makes basic corpus operations like searches and concordances more efficient by disambiguating between different homonymic forms (Atwell 2008: 505). While the annotation of corpus texts with additional linguistic information, like lemma or word class, is considered “important for many purposes” (Johansson 2008: 43), historical corpora are much more rarely so annotated than present-day ones. The amount of analytical annotation in historical corpora is mainly limited by the fact that automated linguistic annotation software have been designed for Present-Day English and cannot cope with extensive spelling variation, while manual analytical annotation of corpora is extremely work-intensive and as such beyond the means of most projects.<sup>38</sup>

Pursuant to the documentary approach and corpus linguistic orientation of

<sup>36</sup> In fact, this type of annotation is possibly the most common one in all kinds of editions, corpora and text archives, and is often the only kind present in minimally annotated digital texts. One possible reason for this is that in electronic textual projects focused on search and analysis of data, the appearance of the text and thus its descriptive annotation is often seen as “less crucial” than its logical organization (McGann 2004: 385).

<sup>37</sup> In the present edition this distinction is perhaps best illustrated by the transcription and annotation of words (described in detail in subsections 10.2.2, 10.2.6 and 11.2.5), which are transcribed and descriptively annotated simply as sequences of letterforms with no regard to the words they form, and divided into lexical words by explicit analytical annotation.

<sup>38</sup> For Early Modern and Late Modern English texts, the *Variant Detector* software <<http://www.comp.lancs.ac.uk/barona/ward2/>> developed by Alistair Baron at the University of Lancaster has achieved quite respectable results in normalising spelling variation to allow the automated annotation of texts. However, the spelling variation in Middle English texts has so far proved to be too extensive and unpredictable for anyone to have produced a similar tool for Middle English (ME) texts. Furthermore, the automated grammatical and syntactic annotation of Old and Middle English would require dedicated analysis algorithms due to the extensive differences in their grammar and syntax, even if spelling variation were to be eliminated. The linguistic annotation included in the present edition, consisting of the regularized spelling and basic part-of-speech information for each word, was done manually by using an alphabetically ordered concordance of all the words in all of the six transcriptions in order to ensure consistent interpretation of all instances of the same word-form.

this thesis and the present edition, the principal responsibility of the editor is here seen to be the production of an accurate digital model of the original document to serve as a linguistic research resource, not the definitive analysis of the *work* or even the *texts* represented by it. This naturally locates the editorial focus on the production of detailed documentary and descriptive layers of annotation, limiting the editorial role as a provider of analytical metadata to the minimum required by the intended use of the edition. In general terms, the editor should provide analytical metadata only on those aspects of the textual object that the intimate knowledge of the edited *documents* has given her special expertise not readily available to other scholars, such as the specific language variant, textual conventions and subject matter of the particular document. All further analytical annotation should be seen to be the responsibility of the scholarly users of the edition—which category can also include the original editor, but in a different capacity, with the annotation also having a different, non-editorial status.

In accordance to this principle, analytical annotation considered to be essential for a corpus linguistic edition like the present one is therefore limited to the following three kinds, listed in decreasing order of essentiality: 1) annotation of the logical structure of the text down to the level of individual word-units, 2) annotation of the lexicogrammatical identity of these word-units, and 3) explanatory notes relating to the more obscure aspects of the subject matter of the work. The first of these is necessary not only for contextualising the linguistic features of the text in corpus linguistic analysis, but also for providing a textual coordinate system to which further overlays of analytical annotation can be mapped. The second type of analytical annotation is intended to provide corpus linguists with an edition approximating the kind of “proper historical corpus” envisioned by Lass (2004b) and described by Keith Williamson<sup>39</sup> as a ‘protean corpus’, where “all words and morphemes are lexico-grammatically tagged, but in a way that is – beyond a traditional surface taxonomy of lexical and morphemic constituents – agnostic with respect to syntactic theory” (Lass 2004b: 40-1), providing a solid basis for more elaborate linguistic analyses of the edited texts. The third kind, on the other hand, is intended to transfer to the user at least some of the specific expertise acquired by the editor through his inevitably intimate acquaintance with the subject matter of the work and the specific idiosyncrasies of the edited documents.

#### 5.6.4 Consequences of layered annotation

As mentioned above, the principal effect of the conceptualization of the data and metadata contained in a digital edition as layers is to highlight the differing ontological status of the descriptive annotation on the one hand and documentary and analytical annotation on the other, the former representing data and the latter two different kinds of metadata. On a more practical level, it serves to define a clear point of demarcation for editorial responsibility, allowing the analytical layer to be opened up to user involvement through the addition of new annotation overlays while preserving the integrity of the data, i.e. transcription and descriptive annotation. This, in turn helps break down the monolithic nature of the edition as an editorial product, allowing it to function as a collaborative research resource

<sup>39</sup> In an abstract for a workshop on the *LAEME* and *LAOS-1* corpora at the 2002 International Conference on English Historical Linguistics 12 conference in Glasgow.

and fostering the accumulation and integration of new metadata, based on the analysis of the existing layers of data and metadata without obscuring them:

This system in which all editorial activity, whether by editor or reader, takes place in overlays that leave information at lower levels undisturbed, encourages the editor to respect the textual evidence, and the reader to respect both the textual evidence and the editor's labor and expertise. (Baker 1998: 271)

### Challenges of multiple annotation layers

The cost of this multilayered and open structure is naturally added complexity and potential for conflict between the different layers of metadata describing the same set of data. As McGann (2004) has observed, all textual objects are organized according to several concurrent structures:

Texts have bibliographical and linguistic structures, and those are riven by other concurrencies: rhetorical, grammatical, metrical, sonic, and referential structures. The more complex the structure, the more concurrencies are set in play. (McGann 2004: 387)

The integration of several ontologically separate layers of information represented by these concurrent layers of annotation is a difficult task for any form of structured annotation. Since the simplest and most effective annotation solutions for any individual aspect of a text—be it visual layout, textual structure or linguistic features—often accomplish their simplicity and efficiency at the cost of precluding—or significantly hindering—the addition of further annotation layers to the text, the annotation solutions adopted in this type of a layered edition must always be a compromise. The core problem raised by this kind of integration of multiple annotation layers is the often discussed but never yet conclusively solved problem of *overlapping hierarchies*, which is a common limitation of all markup languages based on hierarchical tree structures—including XML, the current standard in structural markup languages (see subsection 5.7.1).

While there have been attempts to overcome this problem, either through different kinds of solutions based on existing hierarchical markup languages such as XML, or by defining new markup languages that do not presuppose a hierarchical tree structure, “no current solution combines all the desirable attributes of formal simplicity, capacity to represent all occurring or imaginable kinds of structures, [and] suitability for formal or mechanical validation” (TEI Consortium 2014: 621). The solution adopted in the present edition is twofold and based on defining one method for reconciling the descriptive annotation of the document and the basic analytical annotation of its textual structure within a single source document, and another for the unlimited addition of further analytical data by not only the editor but by any user of the edition.<sup>40</sup>

<sup>40</sup> The first method, described in subsections 5.7.1 and 10.1.4, is based on mechanisms developed for circumventing the strictly hierarchical nature of XML, and has been chosen as a compromise between simplicity of annotation and expressive power, following the recommendations made in the *TEI Guidelines*. The principles of the second method, based on the addition of further analytical metadata in the form of annotation overlays separate from but linked to the data using hyperlinking, is described below, while its technical implementation in the present edition is described in sections 10.3 and 11.9.

### Separation of annotation layers

Some projects are now experimenting with offset markup, arguing that the text file can be kept in a more pristine state (and therefore a more preservable textual object) and that different kinds of markup can then be kept in separate files, all pointing to the original text file. The only markup that the text file would then hold is locational information that the offset files can use to point to the correct parts of the text. (Deegan 2006: 367-368)

As described by Deegan above, ‘offset’ or *stand-off markup* refers to annotation that is not inserted into the data stream at the point it describes (unlike ‘inline markup’), but is placed either elsewhere in the computer file containing the data or in an entirely different file, and connected to the *locus* of annotation through hyperlinking. This idea of separating the annotation (or ‘markup’) from the ‘pristine text’ has recently been proposed by many digital editors as a solution to the complexity caused by multiple annotation layers. In addition to solving the problem of overlapping hierarchies by allowing conflicting annotation structures to be stored separately and applied selectively to the base data according to the current needs of the user (Berrie et al. 2006: 274), the use of stand-off markup has been credited with preserving the stability of the editorially established base text while allowing cumulative user annotation of the text (Eggert 2009). One system, described by Eggert (2009) and called “Just-In-Time Markup (JITM)”, is based on the division of the textual content of the edition from the annotation, or as Eggert characterizes it, on “splitting off text from interpretation” (75). In practice this is accomplished by having a separate “base transcription file”, consisting “only of the verbal text contained within uniquely identified text-element tags”, and any number of “external tagsets”, which the user can use to “present” or “interpret” the base transcription file as she views the text (‘just in time’) (76). The same idea of separating the ‘primary text’ from its annotations using stand-off markup is also implemented in the XML Corpus Encoding Standard (XCES), as pointed out by Lehmborg and Wörner (2008: 489).<sup>41</sup>

While the general approach of using stand-off markup to ensure the stability of texts while allowing collaborative annotation is a commendable solution that is also adopted by the present edition, the aim of separating ‘the text’ from *all* annotation is an untenably naive one from a documentary viewpoint. Especially in the case of medieval manuscript texts, the majority of editorial annotation is in fact descriptive, and thus ontologically parallel to the text itself, either representing textual elements like special characters<sup>42</sup> or describing the layout and structure of the textual stream, including its discontinuities and parallelisms. Separating this

<sup>41</sup> It is unfortunate that the XCES project seems to have flagged somewhat, as it is still based on the seriously outdated P4 version of the TEI Guidelines, there is still no proper documentation to accompany the XML schemas, and the project web site (<<http://www.xces.org/>>) has last been updated in 2008.

<sup>42</sup> Despite its extensive coverage of characters, the Unicode standard—and even its extension by the Medieval Unicode Font Initiative (MUFI) still lacks encodings for a number of symbols required for the transcription of mediaeval manuscripts, especially several of the common abbreviation markers.

kind of descriptive annotation from the text-based transcription of the document is does not represent “splitting off text from interpretation” but rather splitting off one aspect of the data from another. This has not gone unnoticed by Eggert (2009), who also admits that the base transcription files of a JITM system, containing “a single, stable string of characters” (77) cannot be considered to “satisfactorily represent, by themselves, the text of a printed or handwritten document” (80).<sup>43</sup>

For this reason, the present edition does not attempt to separate ‘text’ and ‘annotation’, but rather uses a similar method to separate data from metadata, as defined above. I consider this solution to not only provide a more accurate model of the ontology of the original document as a material artefact, but to be more respectful of the “fundamental distinction [...] between what is physically present in the source and what is not” (Driscoll 2006: 260), explicitly separating editor’s interpretation of what *is* in the document from her interpretation of what it *means*. In practice this means that the ‘base data file’ of the present edition does not constitute a plain text transcription, but rather a descriptively annotated structured transcription, to which annotation overlays containing analytical annotation are attached through hyperlinking.<sup>44</sup> In order to structure the transcription and to provide a ‘textual coordinate system’ (see subsections 10.1.3 and 11.2.5) that allows analytical annotation to be explicitly linked to specific points in the data, the base data file is also analytically annotated with the logical document structure down to the level of individual word-units, each structural element being explicitly identified with a unique identifier.

### Addition of new annotation layers

As Kirschenbaum (2004: 538) has observed, “[m]ost digital humanities scholarship is produced incrementally, in layers” (538). This means that digital editions, being a significant vehicle of digital humanities scholarship, should serve not only as tools for the production of such incremental scholarship, but also as repositories for its results. In this ideal case, the answers to one research question posed to the edition would be fed back into the edition, enriching it and in turn allowing it to answer new kinds of research questions. In order to foster this kind of cumulative and collaborative scholarship, digital editions must—as Eggert (2009: 73) has also argued—consciously allow for and enable the accumulation and integration of collaborative interpretation contributed by their users.

As several scholars have observed (see e.g. Berrie et al. 2006, Lehmberg and Wörner 2008 and Eggert 2009), the most significant problems in the concept of integrating new information to an existing resource are maintaining the integrity of the existing data in the process, and avoiding the uncontrolled proliferation of

<sup>43</sup> Although he, rather oddly, also argues them to be “pleasingly honest in their sparseness” (80), not elaborating any further on what exactly this means. Eggert (2009) excuses the fact that the base transcription file of the JITM system “does not contain the graphic features such as italicizing that we are used to considering as an intrinsic aspect of the text”—or the layout and other graphical features that organize the text into a coherent whole—by saying that the base transcription file is not intended to be “the textual representation that non-expert users will choose to read” (77). This of course forces the question why should markup describing the ‘intrinsic aspects’ of the text not be included in the base transcription?

<sup>44</sup> The documentary annotation pertaining to each transcribed document is contained in a separate ‘header’ linked to but separate from the transcription itself.

different versions of the resource. All of the aforementioned scholars also see the most likely solution to these problems to be the use of independent overlays of analytical annotation in the form of stand-off markup, linked to a stable 'primary text'. In addition to solving the problem of stability, this approach would also mediate cooperative analysis and development of the resource, as layers of stand-off annotation could be independently and concurrently developed by different annotators and selectively applied by users to meet different needs (Berrie et al. 2006: 274). Additionally, it would also be possible to create mutually exclusive annotation overlays, containing for example two competing syntactic analyses of a text based on two different theoretical frameworks. The user could then choose to use either of these analyses as the basis of her own research, or even duplicate the research using each of them in turn and compare the results, gaining insight into the differing implications of the two theoretical frameworks.

In order to answer the call made by Ore (2004: 40-1) and others for an encoding system that would allow users to add new analytical metadata to the edition—and to selectively apply metadata created by other users—the present edition employs the concept of *annotation overlays* that are aligned to the data of the edition using a system of textual coordinates based on the segmentation of the data to uniquely identified word-units (see subsection 10.1.3). An annotation overlay is here defined in general terms as a structured representation of analytical annotation associated with a specific dataset, defining an analytical framework and a series of links between specific *loci* in the annotated data and features of that analytical framework. A simple example of such an overlay would be one which contained a definition for the concept of a sentence in a specific syntactic theory and a series of aggregating links that associated this concept with sequences of uniquely identified word-units in the data, indicating that each such sequence constitutes a sentence in this theoretical framework. A more complex example could not only annotate the sentences but further subdivide them into clauses, define a taxonomy of syntactic functions and link these functions to individual clauses and word-units within the sentences, resulting in syntactically parsed data. A different scholar could then use the first of these overlays to segment the data into sentences and create a further annotation overlay by classifying these sentences into e.g. declarative, interrogative, exclamatory and imperative sentences.

The kind of incremental and cumulative generation of analytical data allowed by this kind of stand-off annotation means that the results of basic analytical tasks frequently performed by linguists or literary scholars—which often involve significant amounts of manual work—could be annotated into the data itself and shared, obviating the need for different scholar to perform these basic tasks over and over again. While it has often been claimed that the detailed analytical annotation of editions is too time-consuming and labor-intensive to be viable, these arguments ignore the fact that regardless of this, scholars working on digital texts do this all the time: for example corpus-linguistic analysis always involves hand-sorting of data located through searches and the use of the researcher's judgement (Curzan and Palmer 2006: 24). The problem is merely that this work is not perceived as a form of annotation, and the raw data resulting from it is discarded after it has served its purpose as the basis of a specific scholarly argument.<sup>45</sup>

<sup>45</sup> A good example of this attitude is provided by Warwick (2004), who argues that it is not worth

The kind of system described here, based on the persistent identification of the elements making up the data would allow users—with suitable tools<sup>46</sup>—to save the results of their analyses for future use with very little additional effort. I argue that this kind of infrastructure for the layered accumulation of analytical data around digital editions does not in fact require much additional effort from the individual editor, but would allow digital editions to fully realize their potential as scholarly work-sites and foster new kinds of collaborative and interdisciplinary scholarship.

## 5.7 The technology: XML and the TEI Guidelines

In order to build a bridge between the theoretical discussion of the ontology of digital editions in this chapter and the more practical discussion of the editorial practices and annotation solutions in chapters 10 and 11, the final section of this chapter will briefly describe the technological solutions employed in the present edition, namely *eXtensible Markup Language* (XML, Yergeau et al. 2008) and the *Text Encoding Initiative's Guidelines for Electronic Text Encoding and Interchange* (*TEI Guidelines*, TEI Consortium 2014). Together these two specifications define the syntax (XML) and semantics (*TEI Guidelines*) of a general purpose markup language for the annotation of digital texts, of which the annotation system described in chapter 11 is a subset defined for a specific purpose.

### 5.7.1 eXtended Markup Language

Nearly all recent and emerging formats and software programs dealing with textual data either exclusively use XML (eXtensible Markup Language), or implement ways of dealing with XML data. The use of XML can be considered the minimum requirement for encoding texts in a sustainable and computer-processable way.

(Lehmberg and Wörner 2008: 485)

As Lehmberg and Wörner above point out, XML has over the last decade become the *de facto* standard not only for the scholarly digitization of texts, but also for web services and office software.<sup>47</sup> XML is a textual data format, with the

---

annotating the figurative uses of words in a text, because “the activity of performing this kind of markup would be so labor-intensive that a critic might just as well read the text in the first place” (371), completely ignoring the fact that unlike merely reading the text, the act of annotating into the text what one has discovered through that reading would result in a permanent enhancement of the text itself, potentially useful for later research.

<sup>46</sup> The development of these kinds of tools for corpus linguistics, the technical requirements for which are relatively simple, is intended as a follow-up project for the present edition, together with the general technical specification for annotation overlays supported by these tools.

<sup>47</sup> For example the current versions of Microsoft Office, Open Office (and its descendant Libre Office), and Apple's iWork all save their data using XML-based document formats, and most web communication formats like Rich Site Summary (RSS), Atom, Simple Object Access Protocol (SOAP), and eXtensible HyperText Markup Language (XHTML) use XML syntax.



data expressed as Unicode-encoded text,<sup>48</sup> embedded within a hierarchical structure of *elements*—consisting of start and end *tags*—that represent various kinds of data entities whose structural properties are described by a *schema*, thus defining a specific XML language. XML itself is thus a metalanguage that specifies a syntax for the construction of various markup languages. Unlike for example HTML, XML thus does not have a closed set of elements, but the schema of each XML language defines the elements used in that language and their structural properties in terms of the syntactic relationships that are allowed between them.<sup>49</sup> The semantic meaning of these elements, however, is not defined by the schema but must be separately documented, usually through a human-readable manual or set of guidelines, such as the TEI Guidelines.

As Fraistat and Jones (2006) put it, the central point in encoding texts using XML is “to think structurally about [...] texts and to allow the computer to think structurally about them as well” (112). In other words, it is a way of making the various textual, paratextual and analytical structures found in a document, as well as their properties explicitly visible not only to the human reader, but also to the computer. One of the greatest benefits of XML as a textual data format—in addition to the fact that it is thus especially well-suited for representing textual data—is that unlike binary data storage formats, it is readable by both humans and computers. However, contrary to what some corpus linguists used to ‘plain-text’ encoded digital texts have argued,<sup>50</sup> XML-encoded texts are not intended to be read as such, but rather presented to the reader through presentation software using various kinds of visual formatting based on the information encoded by the XML elements, thus separating data from its presentation. The status of XML as an Internet standard and the existence of a comprehensive set of standardized, widely supported, and open ancillary technologies like XPath, XPointer, XQuery, XSLT (eXtensible Stylesheet Language Transformations), XSL-FO (eXtensible Stylesheet Language Formatting Objects), XHTML and CSS mean that XML documents can be readily processed by standard web application frameworks and web browsers, which makes the development of both local or online user interfaces using existing Open Source components extremely efficient. Its wide adoption within the digital humanities community—including not only editorial projects but also libraries and archives—also makes it technically easy to exchange data encoded in XML with other projects and thus facilitate the division of “the labor of photographing, cataloging and editing” documents among several parties, each with their own specific expertise.

As was mentioned above, XML is based on a *hierarchical* structure of elements, built on the principle of ‘nesting’ or containment. In addition to this, it is also an *ordered* structure, with elements following each other in a linear order. This particular ontology, which is known as an Ordered Hierarchy of Content Objects (OHCO), has also become the dominant ontology in the field of text encoding and

<sup>48</sup> Unicode is an industry standard for text encoding, developed to “overcome the problems and incompatibilities resulting from the large amount of character encoding schemes that exist, with the aim of representing every writing system in the world” (Lehmberg and Wörner 2008: 485)

<sup>49</sup> For example HTML itself can be expressed using the XML syntax, giving rise to XHTML, which is one example of an XML-based markup language.

<sup>50</sup> Talking of the more generalized superset of XML, SGML, which was the markup language used before the development of XML, Markus (1997) complains that texts “abundantly equipped” with markup like SGML “are only legible by machines, but not by human beings” (217-8).

electronic publishing, which see text as an OHCO (see subsection 2.1.2). While this view offers a very useful conceptualization of textual structure on any one level, e.g. seeing it as a series of chapters, each containing a series of sections, in turn containing paragraphs, or alternately seeing it as a series of pages, each containing a series of lines, it becomes problematic when one tries to conceptualize the text from several different viewpoints (Hockey 2004: 368). Because the OHCO model by definition constitutes a single hierarchy, it is usually used to represent the logical structure of the text as a series of chapters, sections and paragraphs, ignoring the physical context of the document by which it is embodied.

As a result of this ontology, also the XML document is defined by a single *root node*, which contains all the data in the document. This root node contains a sequence of variable number of *child nodes*, which can be either XML elements, *attributes* describing them in some respect, or spans of textual data (*text nodes*).<sup>51</sup> While attributes—which are associated with a corresponding *value*—and text nodes are always what are called terminal or *leaf nodes*, each XML element node can in turn contain an unlimited sequence of child nodes of any type, forming a tree-like hierarchical structure of unlimited depth, in which each node is associated with exactly one *parent node*. This means that each node has a number of *ancestor nodes* equal to its position in the hierarchy (the most distant ancestor of each node being the root node) and each element node has a hierarchical structure of *descendant nodes* including all of its children, all of their children and so on.

### Problem of overlapping hierarchies

The most significant shortcoming of hierarchical markup languages like XML for annotating texts from a documentary point of view is thus their inability to simultaneously represent multiple hierarchies in a single document, resulting from the fact that the OHCO model—and thus XML—does not allow elements to *overlap*. In fact, the OHCO model and XML are perfectly capable of encoding several hierarchies describing different things—after all neither of them are concerned with the semantic aspect of what a hierarchy *means*—as long as they form cleanly nesting subhierarchies of each other. In other words, as long as each node in one hierarchy is contained in its entirety not only by its own parent node but also by a single node of the other hierarchies. This means that an XML document is capable of representing both the logical structure and the physical structure of a text *as long as*: 1) each chapter and section starts at the beginning of a new page, 2) each page contains only whole paragraphs, no paragraph crossing the page boundary, and 3) each paragraph starts at the beginning on a new line. Even in this extremely simplistic example, any real document is likely to break the second constraint, and depending on the length of sections, even the first.

To overcome this problem, markup theorists have developed various ways of representing multiple hierarchies in the same data, either through different kinds of XML-based solutions or by defining new, non-XML markup languages that do not presuppose a hierarchical tree structure. Tennison (2008) summarizes the five common XML-based solutions that have been proposed as follows:

<sup>51</sup> The XML specification (Yergeau et al. 2008) also defines other types of nodes like processing instructions and comments, but since they are not relevant to the overall structure of XML, they are not discussed here.

- a) use of multiple document versions,
- b) replacement of enclosing elements by empty milestone elements,
- c) flattening of the hierarchical structure,
- d) fragmentation of overlapping elements, and
- e) use of standoff markup.

Examples of proposed alternatives or extensions to XML include the Layered Markup and Annotation Language (LMNL) model developed by Jeni Tennison and Wendell Piez (see <<http://www.piez.org/wendell/LMNL/lmnl-page.html>>), the General Ordered-Descendant Directed Acyclic Graph (GODDAG) data structure model and the associated TexMECS notation language developed by the *Markup Languages for Complex Documents* project,<sup>52</sup> and the MECS (Multi-Element Code System) developed for the Wittgenstein Archives Bergen (WAB) project.<sup>53</sup> Unfortunately, tools supporting these languages are either nonexistent or designed for the needs of a specific project (e.g. in the case of MECS), which means that they do not offer a viable solution for the present edition.

The approach taken in this edition for accommodating multiple hierarchies within the edition combines the XML-based solutions *b* and *d*, applied along the lines described in chapter 20 of the *TEI Guidelines* (TEI Consortium 2014: 621–31), with solution *e*, using the former to resolve overlaps within the ‘core’ transcription document and the latter for the annotation overlays linked to this core document (see subsection 10.1.4). This choice of methods has been made with an eye toward a reasonably simple representation of all the structures occurring within the texts included in this edition and expected to occur in historical manuscript texts of similar nature. The issue of conformance to the schema defined by the *TEI Guidelines* has also been considered, and conformant solutions have been preferred over ones requiring extension. This means that the present edition for example follows the *TEI Guidelines* in privileging the textual structure over the physical one by encoding the latter using empty milestone elements.<sup>54</sup> However, some extensions—which will be suggested for inclusion in future versions of the guidelines—have been deemed necessary and implemented in the edition in order to make the annotation more consistent and to simplify processing.<sup>55</sup>

### 5.7.2 The Text Encoding Initiative and its Guidelines

The encoding and annotation of the electronic edition described in this thesis is based on the current version (P5) of the *Guidelines for Electronic Text Encoding and*

<sup>52</sup> <<http://mlcd.blackmesatech.com/mlcd/Research.html>>

<sup>53</sup> For discussions on the various aspects of the problem of overlapping annotation and the proposed solutions, see e.g. Tennison and Piez (2002), Dekhtyar and Iacob (2003), DeRose (2004), Durusau and O'Donnel (2004), Sperberg-McQueen and Huitfeldt (2004), Iacob and Dekhtyar (2005), Sperberg-McQueen (2006 and 2007), Marcoux (2008), Di Iorio, Peroni and Vitali (2009), Portier and Calabretto (2009), Schmidt (2009) and Stührenberg and Jettka (2009). The Wiki page of the Overlap Special Interest Group of the TEI user community (<<http://wiki.tei-c.org/index.php/SIG:Overlap>>) contains a useful summary overview of many of these proposed approaches.

<sup>54</sup> Although this does make the processing of the text on the basis of the physical document structure somewhat more inconvenient (as Pierazzo and Stokes (2010: 399) have pointed out), the dominant role of the textual structure as a textual coordinate system for additional analytical annotation was considered to favour this decision.

<sup>55</sup> This refers mainly to the addition of the @part attribute indicating fragmentation to the <p> element representing a paragraph of prose text (see subsection 11.2.6).

*Interchange* (TEI Consortium 2014), a set of guidelines and document grammars developed and maintained by the Text Encoding Initiative Consortium. While the guidelines are often referred to as ‘the TEI’, the acronym more properly refers to the TEI Consortium (usually referred to as TEI-C for clarity), an international non-profit organization, whose membership is made up by academic institutions, research projects, and individual scholars. As the TEI Consortium website points out, the guidelines “define and document a markup language for representing the structural, renditional, and conceptual features of texts”, and are intended particularly for “the representation of primary source materials for research and analysis” (<<http://www.tei-c.org/Guidelines/>>). Following the fundamental principles of the TEI—known as the Poughkeepsie Principles from the site of its inaugural meeting—the TEI Guidelines are designed to 1) provide a standard format for data interchange, 2) provide guidance for encoding of texts in this format, 3) support the encoding of all kinds of features in all kinds of texts studied by researchers, and 4) be application independent (Vanhoutte 2004: 12). Over the last two decades the TEI Guidelines, the first version (P1) of which was released in the summer of 1990 (Johansson 2008: 47), have gradually become the *de facto* annotation standard for digital editions (Baker 1998: 277; Deegan 2006: 367; Flanders 2006: 145; Pierazzo 2009: 173), and are being used in more than a hundred and fifty national and international projects all over the world (<<http://www.tei-c.org/Activities/Projects/>>).<sup>56</sup> The latest major version, P5, was released in 2007 under an open-source license and has since received maintenance and feature enhancement updates twice a year. Although the encoding described in this thesis was originally based on an earlier version of the P5 Guidelines, it has been updated to conform to version 2.6.0, released in January 2014.<sup>57</sup>

In practice, the guidelines define a large number of XML elements that can be added to an electronic representation of a text to mark features for automated retrieval, processing and analysis by computer software. In addition to these elements and their semantic properties—i.e. what they are intended to represent—the Guidelines also define the specific contexts in which they can be used, i.e. their contextual syntax, and provide information and examples on their interpretation and application. For providing additional information about the phenomena represented by the tags, the Guidelines also define a selection of attributes that can be used to further subcategorize and specify them.<sup>58</sup> From an ontological viewpoint, they essentially define a multi-levelled classification system that seeks to divide the fluid and continuous reality of textual phenomena into a finite number of discrete categories that can be quantified and analysed.

It should, however, be noted, that while the TEI Guidelines provide guidance for the encoding and annotation of text, they do *not*, as a rule, deal with matters of editorial principles or mandate a specific type of edition. In fact, as Rehbein

<sup>56</sup> Even Robinson (2009), who is by no means uncritical of the TEI, freely admits that “without the TEI there would be no Canterbury Tales project and no Wife of Bath’s Project CD-ROM, and any conference on editorial problems focusing on the use of computers would have been very different—or would not have occurred at all” (42).

<sup>57</sup> For people interested in the ideas and history behind the TEI, Vanhoutte (2004) contains a brief introduction to the TEI Guidelines and the TEI Consortium up to that point, and the Guidelines themselves contain a short historical review of the TEI (TEI Consortium 2014: xxiv-xxvi).

<sup>58</sup> However, the permissible values of these attributes are in most cases not defined, or are defined merely in terms of their permitted data type.

(2008) observes, a scholarly edition is just one possible usage of textual data that is encoded on the basis of the TEI schema (7).<sup>59</sup> However, their adoption outside the field of digital editing seems to be more limited. Considering that the TEI Guidelines have—in some form or another—been around since 1990, there are surprisingly few corpora that make use of them, although Lehmberg and Wörner (2008: 485) argue that most of the current standards for the linguistic annotation of corpora can be traced back to the work of TEI or the Expert Advisory Group on Language Engineering Standards (EAGLES).<sup>60</sup> Fortunately, the situation is slowly changing and there are several new corpora that are currently being compiled following the TEI Guidelines, and several historical corpora have also recently been or are being converted into TEI-compliant XML.<sup>61</sup> It should also be noted that the guidelines are not tied to any particular markup technology, although they are currently expressed in terms of XML, defined by the World Wide Web Consortium's XML Recommendation (Yergeau et al. 2008). This means that in the event that a new markup language, more suitable to the purposes of the TEI is developed, the guidelines can be easily translated to that language.<sup>62</sup>

As a result of their stated aim of allowing the encoding of all kinds of texts, the TEI Guidelines are often criticized for being intimidatingly extensive—in book form, the P5 guidelines run to some 1,500 pages—and for providing too much leeway for alternative encodings, leading to inconsistent results that make validation and processing difficult (Huitfeldt 2006: 187; Lehmberg and Wörner 2008: 487). What people sometimes fail to understand about the TEI Guidelines is that they are exactly that: “guidelines that may be applicable as they stand for some projects, but that can also act as a base on which individual encoding schemes can be built for particular projects” (Hockey 2004: 367). As for example Durusau (2006: 299); Lavagnino (2006: 335) observe, and even the guidelines themselves state (TEI Consortium 2014: 668), the guidelines are not really intended to be used as they stand in their entirety, but rather to serve as a selection of encoding solutions from which an individual editor or project can choose the ones most suitable for the task at hand, documenting their choices to create a customization of the

<sup>59</sup> In fact, this thesis has itself been entirely encoded as an XML document following the TEI Guidelines, which has then been transformed into the  $\text{\LaTeX}$  typesetting language (`<http://latex-project.org/>`) using XSLT.

<sup>60</sup> Of the few corpora encoded according to the TEI Guidelines, certainly the most well known is the British National Corpus. Of historical English corpora only the *Lampeter Corpus* (Hofland, Lindebjerg and Thunestvedt 1999), which follows the P3 version of the guidelines, and the *Corpus of Middle English Prose and Verse* were originally encoded using the TEI Guidelines. Unfortunately the latter is only accessible through an online interface which does not reveal the TEI XML encoded source files.

<sup>61</sup> New TEI-based historical corpora include for example the third part of the *Corpus of Early English Medical Writing (Late Modern English Medical Texts)* being compiled at the University of Helsinki and the *The Coruña Corpus of English Scientific Writing* being compiled at the Universidade da Coruña. Converted corpora include the *Dictionary of Old English Corpus* which was converted to TEI P5 in 2009 by Antonette diPaolo Healey, Joan Holland, David McDougall, and Ian McDougall, the pioneering *Helsinki Corpus* that was released as a limited edition TEI XML version in connection with its 20 year anniversary (HC TEI XML 2011), and the *ARCHER* corpus, which is being converted into XML under the coordination of David Denison and Nuria Yáñez-Bouza (`<http://www.alc.manchester.ac.uk/subjects/lel/research/projects/archer/archer-versions/archer3_2/>`).

<sup>62</sup> This has in fact already happened once; earlier versions of the TEI Guidelines, up to P3 were expressed using SGML, the update from P3 to P4 involving their translation to the syntax of XML.

guidelines. One of the principal aims—if not *the* principal aim—of this thesis is to define such a customization for creating corpus linguistic editions, and to test it by creating such an edition. While this chapter has focused on the theoretical underpinnings of such an edition and chapter 10 will describe the more traditional editorial principles followed in creating such an edition, chapter 11 will document the ways in which the TEI Guidelines should be applied for this purpose (and have been applied in the present edition), delineating the “features of the text(s) to be encoded, elements to be used for such features, attributes for the elements used, and the range of values for each attribute that is allowed for each element ” (Durusau 2006: 299).

### Benefits and limitations of the TEI Guidelines

The single greatest benefit of the TEI Guidelines is that they exist, and thus obviate the need to start the design of an annotation system from scratch. As Mahoney (2006)—discussing the digital encoding of inscriptions—points out, “TEI has already addressed many of the taxonomic and semantic challenges faced by epigraphers”—as well as other editors—and furthermore, the “TEI-using community can provide a wide range of best-practice examples and guiding examples” (233-4). From the point of view of the present edition, its documentary nature, and its New Philological and linguistic orientation, it is fortunate that while the TEI Guidelines are not tied to specific editorial principles, the practical guidelines provided do nevertheless imply an essentially documentary approach to editing to the extent that following the recommended practices will result in a documentary edition that fulfils *The Minimum Standards for Electronic Editions* outlined by the Association for Documentary Editing (ADE Committee 2002).

Furthermore, an editor editing for the purposes of corpus linguistics is in a fortunate position, since the primary approach for which the guidelines have been designed—“to encode texts, perhaps also preserving some features of the original document but at a secondary level” (Pierazzo and Stokes 2010: 401)—is especially well-suited to the primarily—but not exclusively—textual focus of corpus linguistics. However, in contrast to traditional approaches to corpus encoding, the TEI Guidelines do not limit the modelling of the textual object just to the textual level but support its encoding on several conceptually different levels, as a physical, typographical and linguistic object (Pierazzo 2011: 467-8; Lehmberg and Wörner 2008: 486).<sup>63</sup>

An even more significant benefit of the TEI Guidelines is the fact that they establish a standard way of modelling the document, obviating the necessity of every individual project to invent “codes, systems or symbols of their own” to indicate documentary features that could not be represented by a text transcription (Deegan and Tanner 2004: 493-4). This kind of standardization, as has been previously mentioned, has several advantages. First of all, it facilitates the creation

<sup>63</sup> While traditional linear text-based annotation systems—like the ones used in the *HC* or the *EWD*—are able to represent many manuscript features like cancelled words, additions written above the line or in the margin, and abbreviation markers to some degree of detail, they cannot readily represent ‘multilayered’ or ‘parallel’ phenomena like cases where one word has been replaced by another by overwriting it, or simultaneously encode both the abbreviation marker actually present in the text and its editorial expansion.

of interoperable resources—which is especially important for editions that are intended to be integrated into a corpus. Second, it establishes a common technical framework “comprehensible to a large body of users and scholars” (Pierazzo 2009: 173), thus fostering co-operation between editors and scholars. Thirdly, it defines a standard, platform independent encoding format, facilitating the preservation of editions in digital libraries, helping them survive changes in computer systems, and fostering the development of software support (Robinson 1998: 254; Crane 2006: 278).<sup>64</sup> While providing stability that can help editions weather technological changes, a standard that is itself immutable will sooner or later become more of a hindrance than an aid as the needs of editors and users outgrow its possibilities (Deegan 2006: 367). In order to avoid this, the TEI has been built as a modular system, with an inbuilt mechanism for implementing and documenting additions and customizations, that can in time be incorporated into more general use. This allows it to both provide the inertia necessary for stability and to gradually adapt to new requirements.

Although some scholars like Robinson (2009: 50) disregard the standardising function of the TEI Guidelines and see them merely as a convenient source for solutions for furthering the editor’s vision of how the text should be presented to the user, I see the standard established by the TEI to have value in itself. As Cummings (2009: 316) reminds us, it is the adherence to open standards that enables the interoperability between different digital resources, and deviating from these standards—even for seemingly good reasons—always poses a risk to this interoperability. Thus, although following the guidelines over an ad-hoc solution might require a more complicated processing implementation, or result in a nominally more complex annotation, I believe these difficulties to be more than offset by the benefits of datasets (i.e. editions) and tools that can interoperate through a widely recognized standard interface like the one defined by the TEI Guidelines.

Despite the numerous advantages resulting from the status of the TEI Guidelines as a *de facto* standard, they are not without their limitations and problems. First of all, they are—as mentioned above—quite extensive and can seem overwhelming when approached for the first time. However, they will feel much less so as one realizes that due to their modular nature, an editor can most likely focus on just the few modules that are relevant to her edition, drastically reducing the amount of information that needs to be learned. In addition to the Guidelines themselves, there are also extensive online teaching materials for learning TEI on your own, including the extremely useful *TEI by Example* series of tutorials, created by leading TEI experts coordinated by Edward Vanhoutte and Melissa Terras (<<http://www.teibyexample.org/>>).<sup>65</sup> The TEI has also been criticized for its focus on “Jesuitical discussions on the minutiae of text encoding and character sets” (Robinson 2009: 43-5). While some of the criticism is justified, Robinson (2009: 43) does quite aptly point out that many of the issues dealt with by the TEI to what

<sup>64</sup> For example the oXygen XML editor (<<http://www.oxygenxml.com/>>) already has built-in support for the TEI schemas and can thus assist the user in writing valid TEI XML. And fortunately for corpus linguists, the linguistic analysis toolkit—combining two powerful open-source tools, the that is being developed by the Textométrie project at ENS de Lyon is also designed to work with TEI-encoded documents.

<sup>65</sup> The University of Oxford IT Services also organizes an intensive introduction course to TEI each summer, the TEI@Oxford Summer School, <<http://tei.oucs.ox.ac.uk/>>.

seems like infuriatingly excessive detail are far from being as clear cut as some of the critics<sup>66</sup> would like to think.

A second problem often ascribed to the TEI Guidelines is its unsuitability for initial transcription work. For example Robinson (2006) recounts that while the Canterbury Tales project decided very early on to use TEI SGML/XML as the eventual storage format, the working transcripts themselves were not encoded in SGML, since the SGML editors of the time were not easy to use, and the transcription stage required a more “efficient and focused system” (Robinson 2006: 79–80). While the initial transcription for the current edition was not done in XML,<sup>67</sup> my current solution for initial transcription would be to encode the text directly into TEI XML using a dedicated XML editor like oXygen XML Editor, possibly with a visually formatted ‘Author’ mode with appropriate stylesheets. While Lavagnino (2006) considers “the precision and formality required for TEI-encoded texts” as a hindrance at the early stage of editing when the editor “may be entertaining many conflicting ideas about what sort of information will be in their edition and how it will be structured” (336), I argue that the different levels of detail in annotation allowed by the TEI Guidelines also facilitate exploratory annotation, where one initially annotates merely the visual features of the original document and later proceeds to add more analytical annotation based on the analysis of those visual features. Naturally, the optimal transcription situation involves a ready XML schema based on a predefined subset of the TEI which allows a schema-aware XML editor to guide the transcriber into choosing the appropriate elements, attributes and attribute values and prevents her from using syntactically or contextually—though not semantically—inappropriate annotation.

### Related TEI customizations

As was mentioned above, the TEI Guidelines are intended to serve as a basis for defining more strictly limited subsets intended for specific kinds of documents and/or for specific purposes. One of the earliest special-purpose customizations was actually developed for encoding linguistic corpora for the purposes of language engineering. The Corpus Encoding Standard (CES), now called XCES after its shift from SGML to XML with TEI P4) was developed by the EAGLES group in 1996 to address some of the shortcomings of TEI from the point of view of corpus linguists and to provide a more strictly defined exchange format to mediate between various corpus annotation standards (Lehmberg and Wörner 2008: 488). While the aims and general principles are in some respects quite similar to the guidelines presented in this thesis, it is incompatible with the New Philological view adopted here as it entirely ignores the documentary aspect of the textual object and simplifies its textual content into a linear character sequence. Furthermore, this standard seems to have never achieved widespread popularity, and its

<sup>66</sup> Robinson mentions Jeremy Sinclair as an example and I have also personally encountered my fair share of these.

<sup>67</sup> The initial transcription and proofreading of the present edition used a hybrid solution of ad-hoc shorthands and hidden pseudo-XML code created using a custom tool built on OpenOffice macros, which caused a large amount of extra work when the transcriptions were finally converted to TEI XML.



development seems to have stopped after initial draft versions.<sup>68</sup>

Another early customization especially relevant for the present edition in terms of its purpose was developed by the Model Editions Partnership (MEP) for encoding editions of historical documents for historians in a way that would fulfil the scholarly standards required of printed editions. The project produced a draft version of a very detailed set of guidelines, based on the P4 version of the TEI Guidelines, and nine experimental mini-editions that served as test cases for the guidelines.<sup>69</sup> This customization is similar to the present guidelines in its level of detail, but differs from it by being defined primarily by the type of material it is intended for, namely personal and administrative documents such as letters, journals and minutes, and only secondarily by its aim, which is distinctly historiographical.

Yet another relatively early customization, designed for the encoding of correspondence material, is the *DALF*, which is also based on the P4 version of the TEI Guidelines. The *DALF* guidelines make many extensions to the TEI guidelines, motivated by the specialized nature of the texts it encodes; for example the meta-data header is reorganized and redefined to suit the description of modern letters, and a large number of new elements are introduced to describe the specific parts of letters.<sup>70</sup> The *DALF* guidelines are very similar to the present customization not only in being a strictly defined subset of the TEI, defining the syntax, elements, attributes, and attribute values that are available, but also in paying considerable attention to the physical, documentary aspects of the textual objects and encoding the physical structure of the document in addition to its logical structure.

An example of a very well-documented and actively developed TEI customization is provided by the EpiDoc Guidelines<sup>71</sup> for the encoding of ancient documents, which are the main product of the EpiDoc Collaborative (Mahoney 2006: 234). While being concerned mainly with inscribed documents—although papyri are also covered—the EpiDoc customization is very similar to the present one in terms of its documentary focus that takes into account and attempts to encode also the material aspects of the textual object. Another relatively recent and actively maintained TEI customization, based on the P5 version of TEI, is the *Best Practices for TEI in Libraries*<sup>72</sup>, a “guide for mass digitization, automated workflows, and promotion of interoperability with XML using the TEI”. While the purpose and general approach of this TEI customization is very different from that of the present guidelines, it is very similar in terms of the level of detail of its documentation.

Perhaps the most relevant of currently existing TEI customizations is the one

<sup>68</sup> While the XCES web site states that it is “continually under development” (<<http://www.xces.org/>>), the site has not been updated since 2008 and the CES guidelines themselves have last been updated in 1999, the document being described as “a first draft of the standard” (<<http://www.cs.vassar.edu/CES/CES1-0.html>>).

<sup>69</sup> The guidelines have not been updated since 2002, and they are still in a very provisional state but do document the development process of the guidelines in great detail with numerous open questions and musings by the editors. The original web site of the project has disappeared, but it has been archived at <<http://wyatt.elasticbeanstalk.com/mep/>>.

<sup>70</sup> Many of the elements introduced by *DALF*, or close parallels to them, have since been introduced to the TEI Guidelines, and the current P5 version of the guidelines would require much less customization for the same kind of material.

<sup>71</sup> <<http://www.stoa.org/epidoc/gl/latest/>>

<sup>72</sup> <<http://www.tei-c.org/SIG/Libraries/teiinlibraries/>>

developed by the *Medieval Nordic Text Archive (MENOTA)*<sup>73</sup> project for encoding medieval Nordic manuscript texts with an emphasis on their linguistic research. The *MENOTA* guidelines (Haugen 2008), originally based on TEI P4 and subsequently migrated to P5, contain precise instructions on how to encode the specific phenomena occurring in these manuscripts with the aim of producing more uniformly encoded editions (Lehmberg and Wörner 2008: 487). Like the current edition, the *MENOTA* project strictly separates data from its presentation by maintaining the archival versions of their transcriptions—encoded as XML documents—separately from the presentation versions—formatted as HTML or Portable Document Format (PDF) documents—that are generated from the archive versions by XSLT stylesheets (Haugen 2004: 73, 87). The main differences between the approach outlined here and that of *MENOTA* are that the present approach pays more attention to the encoding of the visual and physical features of the document and is stricter in its adherence to canonical TEI solutions, while *MENOTA* defines a relatively large number of project-specific elements and annotation structures. Furthermore, unlike the *MENOTA* transcription system, as described by Haugen (2004), the present edition does not regularize the punctuation or other paratextual elements of the original document, mainly because it is not considered to fall within the sphere of ‘formal regularization’, but rather as constituting a structural alteration of the text.<sup>74</sup>

The considerable similarities between the aims of this thesis and the *MENOTA* project naturally raise the question why does the present edition not adopt the *MENOTA* guidelines instead of defining yet another customization of the TEI Guidelines. There are in fact several answers to this question, some of them historical, some practical and some more theoretical and/or ideological:

- When work on the current edition and its guidelines was begun in 2006, the current version of the *MENOTA* Guidelines were still based on TEI P4, and even though they were nominally migrated to P5 in 2008, they still retain custom encodings that parallel native TEI solutions introduced by P5.<sup>75</sup>
- The three levels of transcription defined by *MENOTA*—‘facsimile’, ‘diplomatic’ and ‘normalized’—were not only considered to be excessive for Middle English (and even more so for later) texts, mainly because of the problematic nature of ‘facsimile’ or *graphetic* transcription (see subsections 5.4.1 and 10.2.2), but also to conflate descriptive and analytical annotation in a way that was considered inappropriate and impractical in terms of the annotation model outlined above in section 5.6.<sup>76</sup>

<sup>73</sup> <<http://www.menota.org/>>

<sup>74</sup> The only reason I can see for the insertion of ‘normalized’, i.e. modern punctuation would be to present an interpretation of the syntactic structure of the text, which is better accomplished formally through the use of the XML elements defined for this purpose by the *TEI Guidelines*.

<sup>75</sup> Examples of this are the <me:punct> element for punctuation symbols, for which TEI P5 introduced the equivalent <pc> (‘punctuation character’) element, and the @me:msa (‘morphosyntactical analysis’) attribute, which duplicates the function of the @ana attribute in TEI P5. These kinds of unnecessary deviations from the TEI standard are problematic in terms of interoperability and are avoided in the present guidelines.

<sup>76</sup> While the *MENOTA* guidelines do not require transcription on all three levels, the three-tiered transcription has led them to again create custom elements instead of using the method defined in the TEI P5, which provides for two levels of transcription (diplomatic and normalized). Furthermore, only using ‘diplomatic’ and ‘normalized’ levels of transcription would have precluded the annotation of abbreviated words, which are here considered extremely important for the study of mediae-

- Despite being designed to produce more uniformly encoded editions, the *MENOTA* guidelines nevertheless contain several alternative encodings for some features without clearly specifying the conditions in which each encoding is to be used,<sup>77</sup> making them completely interchangeable and thus redundant. This is problematic, as it makes the implementation of processing software unnecessarily complicated and creates unnecessary variation between supposedly similar editions.
- Despite the fact that they are not supported by XML Schema, *MENOTA* still uses *entity references*—inherited from SGML—for encoding non-standard characters and symbols instead of the <g> element provided by TEI P5, which means that the *MENOTA* guidelines cannot be fully expressed as an XML Schema but need to rely on the older, more restricted Document Type Definition (DTD) format for their formal specification.<sup>78</sup>

It should, however, be noted that the significant similarities between the two sets of guidelines mean that it is possible to define a relatively simple transformation from the present encoding system to the *MENOTA* scheme without losing any information preserved by its diplomatic transcription level. Since the present scheme encodes much information that is not captured by the *MENOTA* scheme, transformation in the other direction will in most cases require at least some manual work and most likely also consultation of the original manuscript.

## 5.8 Conclusion

At present, XML offers a way forward, enabling more complex document models to be delivered over the Web. In the meantime, I would like to reiterate my plea for more basic research and development on what an electronic edition should look like and on how it can best maintain and enhance current standards of editorial scholarship.  
(Hockey 2004: 374)

In answer to the plea by Hockey and others, the principal purpose of the present edition and the description of its editorial principles and encoding and annotation practices in chapters 10 and 11 is to define a new type of digital edition that allows the linguistic, textual and paratextual features of the original documents to be analysed using corpus linguistic methodology. In terms of its functional ontology, such a digital edition is here seen as consisting of three separate layers: *data archive*, *processing engine* and *editorial output*. Of these, the identity of the edition as an edition of something lies in the data archive, while the processing engine defines the edition in terms of its ‘kind’. The editorial output does not have an independent existence, but is rather the result of the interaction of

---

val textuality (see subsection 10.2.4).

<sup>77</sup> These include for example overlapping structures, described in section 4.10 of the *MENOTA* guidelines, and the encoding of special characters, described in chapter 5 of the guidelines.

<sup>78</sup> This is especially problematic, since the DTD does not support the use of separate XML *namespaces*, which are used to indicate custom elements defined by the *MENOTA*.

the processing engine on the data archive, resulting in a specific view on the data. While the data archive is here seen to contain the essence of the digital edition—an analytical model of the edited document—it is the separation of the data and the editorial output by the processing engine that provides the digital edition with capabilities exceeding those of a printed edition. This separation of data and its presentation is thus a crucial feature of digital editions, since it lies at the root of most of its benefits. By dynamically filtering the data archive according to the user's requirements, it makes possible the inclusion of an unlimited amount of data in the archive without overwhelming the user and making the edition unusable, and also allows the same data to be presented in a multitude of different ways for different purposes.

Most of the requirements posed by corpus linguistic research, outlined in chapter 4, are fulfilled by the use of annotation to represent various logical, textual and paratextual features of the document in the data archive in a structured and machine-readable way. An annotated transcription serving as a data archive for a digital edition is here seen as a model of the original document. This model consists of essentially two components, *data* and *metadata*. Data is made up of the textual transcription and descriptive annotation of the material paratext of the original document, including its layout, visual decoration and emphasis, corrections and additions, and changes in hand. Metadata is here seen to consist of documentary annotation describing both the historical, codicological and palaeographical properties of the original document, and the relationship of the edition to the original (i.e. the editorial principles and practices employed), and analytical annotation describing individual parts of the textual object embodied by the document with respect to an analytical framework of some kind. All of these layers of annotation can also be seen as having clearly differentiated functions: *documentary annotation* tells us what the record represents, while *descriptive annotation* tells us what the record is like in terms of its perceptible features, and finally, *analytical annotation* places the linguistic event reconstructed from the text and the preceding annotation into the context of some theoretical framework, allowing us access to patterns and meanings that are not visible or immediately obvious in the original record.

This separation of the different layers of annotation allows for a sensible division of labour between the different people involved in the process of (digital) humanities scholarship, based on their specific expertise. In the optimal case, the descriptive annotation would be prepared by a librarian familiar with the bibliographic features and provenance history of the original document, as well as its relationship to other related documents. The descriptive annotation, along with the interpretation and encoding of the textual content of the original document, would fall into the province of the philologist editor as an expert on the particular document and its material, social and cultural contexts of production. Finally, the analytical annotation would be produced jointly by a variety of scholars, each placing the textual object modeled by the edition within a particular theoretical framework and reporting the results of their analysis in the form of new analytical annotation overlays.<sup>79</sup> From a technological standpoint, all of these layers of

<sup>79</sup> It should be noted that this division of labour is based on scholarly roles and not necessarily specific individuals—a single individual could very well serve as an archivist, editor and scholar for a specific

annotation are realized using a restrictively customized version of XML markup language defined by the TEI, which allows all of the annotated features to be processed and analysed by computational means and related to the textual content of the edition, allowing them to form a para- and extralinguistic context for the corpus-linguistic analysis of linguistic features. This can allow us to discover and make explicit things which were implicit or invisible in the original record, and thus to get more out of our research materials.

Traditionally, linguistic corpora have been reasonably well furnished with documentary annotation and in the case of modern language corpora, often also with analytical annotation. However, historical corpora have been worse off in this regard, since various factors like spelling variation have made the automatic analysis and analytical annotation of historical texts difficult. Descriptive annotation of linguistic corpora, on the other hand, is usually very sparse, if it exists at all,<sup>80</sup> being often seen as interfering with the linguistic analysis and the addition of analytical annotation to the text. The purpose of the editorial approach and annotation scheme presented in this thesis is to develop a way of fruitfully combining descriptive and analytical annotation in a way that allows them to support rather than detract from each other.

The principal method for this is the creation of a textual coordinate system that allows overlays of analytical stand-off annotation to be persistently linked to the original editorial data archive without the need to make changes to it.<sup>81</sup> This allows linguist users to analyse the annotated text and to attach the results of their analyses to the original data without disturbing it. Furthermore, it also allows for the inclusion of a multiplicity of different analytical layers without needing to worry about their structures conflicting with those of the descriptive annotation or of each other. This allows both the editor and the linguistic scholar to concentrate on their own areas of expertise and to create something useful together.

In an ideal case this will lead to an accumulation of knowledge, as people share their work on a given document and build on each other's work. This, in turn, means that one kind of analysis only needs to be done once—if done properly—eliminating the need for everyone to perform the same basic research tasks over and over and allowing them to move on to more advanced research questions, standing on the shoulders of their predecessors. As Cummings (2009: 310) has pointed out, this kind of an *agile edition*, consisting of “texts with fairly fine granularity of structural markup in a main file but with additional resources and markup provided in a stand-off manner in other files”, would be easily reusable for different research purposes and amenable to further annotation by others and would “allow for a greater degree of flexibility in their later uses”, which is exactly what is required in both historical corpus linguistics and the humanities in general.

---

document, but she could still be seen as performing these different stages of the editorial process in different capacities.

<sup>80</sup> Descriptive annotation seems to be traditionally associated more with literary or historical digital editions, which do not often fulfil the needs of corpus linguistics described above.

<sup>81</sup> The Textométrie project at the ENS de Lyon (<<http://textometrie.ens-lyon.fr/>>) uses a similar technical solution for their open-source tools, which will most likely make these tools a good starting point for the development of a full-featured online analysis interface for the present edition and others following similar guidelines.



## **Part II**

# **Historical background**





## Chapter 6

# Linguistic context: multilingualism and literacy

Considering the linguistic orientation of the present edition, an important aspect of the context in which texts were produced and consumed is their linguistic environment. As was observed in section 2.2, the linguistic context of textual production affects not only the specific linguistic codes that are available for encoding ideas and meanings into *text*, but through its interaction with the surrounding cultural context, also the variety and characteristics of discourse types, genres and registers that are used to convey different types of ideas. In more practical terms the (socio-)linguistic context determines what linguistic resources are available to language users and how these resources depend on their sociocultural status. The late Middle Ages were a linguistically interesting period: the process of vernacularization had significantly expanded the purview of written Middle English after a long period of mainly Latin and Anglo-French literary culture, and the spread of literacy down the social scale and the increasing prominence of the rising middle classes were creating new markets for written texts, which would ultimately be catered for by the emergence of printing at the end of the 15<sup>th</sup> century. Since understanding the social and cultural associations of the different linguistic resources available to language users in late medieval England is crucial for understanding the significance of linguistic choices made by speakers and writers, this section will briefly outline the linguistic situation of late medieval England, including its fluid and changing multilingual nature, the gradual emergence of a more standardized form of Middle English through the influence of the metropolitan culture of London, and the extent and nature of literacy among different parts of the population.

### 6.1 The trilingual environment of medieval England

The linguistic repertoire of late medieval England was complex, unstable, and socially charged. If the languages an individual used—Latin, French, English, or any of the indigenous Celtic languages—

were in part functions of birth and upbringing, their use in particular domains helped sustain the dynamics of society. Like individual speech acts, moreover, languages had meaning in relation to one another. (Machan 2009: 363)

As the result of complex historical developments—in simplistic terms the Christianization of Europe and the Norman Conquest of 1066—medieval England was effectively trilingual between English, Anglo-Norman French and Latin. While the roles and functions of the three languages changed significantly from the 11<sup>th</sup> to the 15<sup>th</sup> century, the basic pattern remained similar: Latin was the language of religion, culture and power, institutionalized throughout Europe, French had been introduced as a language of the ruling class as the result of a violent historical event “that turned upside down the political structure of the country”, and English had been ousted from many functions by the two more prestigious languages, although “never completely eradicated” (Crespo Garcia 2000: 23). This basic situation has been summarized by Crespo Garcia (2000: 24) as shown in Table 6.1.

The roles of the two extremes, Latin and English, were always relatively clear: Latin was “the language of tradition, authority and power”, while English was “the language of the people, impermanence, and change” (Machan 1994: 145). The role of French—or *Anglo-Norman* as the variety spoken in medieval England has been called—was more ambiguous. In the later Middle Ages it was neither the true vernacular of even the English nobility, nor was it a fully established and formally codified language of record like Latin. However, as Clark (1992) has pointed out, the trilingual situation in 13<sup>th</sup>- and 14<sup>th</sup>-century England nevertheless involved “not two vernaculars and one learned language, but one vernacular and two learned ones” (125-6). While the Norman Conquest had displaced English as a language of official documents, English still remained not only the spoken language of most of the population but was also employed as a language of “teaching and preaching, and of disseminating general didactic and encyclopaedic knowledge”, especially after 1100 (Treharne 2011: 220). Furthermore, as Smith (2008: 215) points out, while literacy in English seems to have been on the increase during most of the Middle English period, Middle English was predominantly *local* and parochial in its functions until the very end of the period, Latin and French being adopted whenever writing was needed for national purposes. This contextually contingent multilingualism has prompted Machan (1994: 3) to argue that the term *Middle English* should be used to refer to “*how* the language was used during the period”, not just to recognize “that specific linguistic forms were utilized in specific works during a specific period”:

If a term like *Middle English* is to have broad cultural applicability, it should also account for which speakers were using which forms of the language in which social contexts and for which culturally conditioned reasons. [...] To say that a text is in Middle English, therefore, is to impute to it not only a range of linguistic and literary forms but also a range of sociolinguistic contexts in which it meant, as well as a variety of cultural practices it mediated. (Machan 1994: 3)

LANGUAGE	Register	Medium	Status
Latin	Formal-Official	Written	High
French	Formal-Official	Written/Spoken	High
English	Informal-Colloquial	Spoken	Low

**Table 6.1:** *Relationships of the languages used in medieval England (after Crespo Garcia 2000: 24).*

In this vein, the characterization of the *Potage Dyvers* and other medieval English recipe collections as “Middle English” is here intended to highlight their intimate connection with a specifically *English* culinary culture and tradition (see chapter 8) rather than any strictly linguistic classification; after all, the tradition of English culinary writing was heavily indebted to its French counterpart and made extensive use of French and Anglo-Norman lexis and even syntax, as is obvious also in the present edition.

The fact that medieval Britain was effectively trilingual for much of the Middle Ages does not mean that all individuals knew several languages; the greater part of the rural areas and much of the lower classes were most likely monolingual English-speakers, while many members of the highest nobility of the early Middle English period were most likely effectively monolingual French-speakers (Schendl 2000: 77). While “by no means an unusual achievement with the literate part of society” (Schendl 2001: 310), bi- or even trilingualism—in both written and spoken registers—seems to have been most common among those members of the middle ranks of society who would have dealt with both the upper and lower extremes of society (77), such as “parish priests, merchants, country stewards and wet-nurses” (Nevalainen and Tieken-Boon van Ostade 2006: 272-3). This kind of linguistic stratification is also echoed in the well-known contemporary assessment of the linguistic situation in 14<sup>th</sup>-century England by Robert of Gloucester, who saw French and English “as essentially sociolects, the former spoken by the ‘heie-man of þis lond’ and the latter by ‘lowe men’” (Machan 2009: 367). As Nevalainen and Tieken-Boon van Ostade (2006) observe, the functional division of the three languages after the Norman Conquest also corresponded to the medieval division of the society into the three estates: “those who normally fought used French, those who worked, English, and those who prayed, Latin” (273).

This hierarchical distribution of languages was also reflected in patterns of textual transmission, texts being mostly translated from high-prestige languages i.e., Latin, Greek and French, the aim being not so much to conform the text to the target language, but rather to elevate the target language to the level of the source language in terms of its expressiveness (Blake 1992b: 7–8). This naturally resulted in significant amounts of lexical and even morphological influence on English, as will be shown in section 6.3. However, as Treharne (2011: 224) observes, the transmission of texts between the three languages of medieval England was not unidirectional with original material written in Latin and then translated into Anglo-Norman for the upper classes and into English for the lower classes, but

English texts were also occasionally translated into French and even Latin.<sup>1</sup>

### 6.1.1 Changing relationships of England's languages

In addition to being functionally and contextually distributed, the respective roles of the three languages also changed throughout the Middle Ages, especially with regard to their functions and domains: while French became “increasingly restricted to a small number of functions such as law in the late ME period” and English “extended its functional range”, Latin “maintained its status as the High variety in most functions throughout the ME period” (Schendl 2000: 77-8). All three languages seem to have been “equally viable choices for writing in the two centuries after the Norman Conquest of 1066” (Treharne 2011: 220), being used for literary texts in all the major genres (such as historiography, hagiography, homilies, sapiential writing and pedagogic texts). However, there seems to have been little integration or interaction between them, as for example the English and Anglo-Norman literary traditions seem to have existed strictly separately, not having any contact or occurring together in manuscript compilations, but rather forming “separate discourses, possibly for quite different audiences, and emerging from distinct traditions” (Treharne 2011: 221-2). By 1400, the use of English as a spoken language and Latin and French in official, administrative writing had established “a clear dichotomy between the colloquial language and the official written language” (Fisher 1977: 874), which Fisher sees as paving the way for the creation of written standard English in the early 15<sup>th</sup> century (874). During the same period, the written language use of the professional middle classes, as reflected for example in the archival records of the Goldsmiths' Company, seems to have made use of both these registers, as it exhibits “a completely trilingual situation” well into the 15<sup>th</sup> century (Jefferson 2000: 183).

Although the changes in the English language situation have often been seen as “a simple concerted struggle between English and Latin and French” Machan (2009) has argued against this, pointing out that it should more accurately be described as “a more general reconfiguration of the repertoire of English and England”, driven by developments like “the expansion of markets, the growth of printing and literacy, political centralization, the spread of schools, colonial aspirations, and nationalistic consolidation” (372). This view is also supported by the mixed language of English culinary writing, as well as by the extensive survival of macaronic lyrics in French and English from the late thirteenth and early fourteenth centuries, which Treharne (2011) sees as an indication that instead of competing for space, English and Anglo-Norman were in fact “comfortably accommodating” of each other (Treharne 2011: 235). This meant that for example scribes writing professionally in a variety of genres were trilingual on a very practical level; as the pragmatic function of each text determined its language and textual conventions, scribes would constantly switch from one language and style to another in the course of the day, creating an environment that was extremely conducive to interaction between the languages (Schendl and Wright 2011: 21).

<sup>1</sup> As an example of this, Treharne mentions the *Ancrene Wisse* (or *Ancrene Riwe*), which was written in English, copied widely through the 13<sup>th</sup> to the 15<sup>th</sup> centuries with various adaptations, and also translated into French and Latin.

Hunt (2000) has argued that the kind of multilingualism that obtained in medieval England is not in any way unusual, but actually the norm in every society “outside a few western societies with a strong sense of language identity and near-universal literacy” (Hunt 2000: 131). Furthermore, Treharne (2011) has concluded that the textual universe of medieval England was in fact more fluid and less linguistically segregated than is often assumed by our scholarly dichotomies of “secular versus religious, French versus English, educated versus uneducated, written versus oral, central versus marginal”, arguing that “if we are to understand a complex era of strategic literacy, generic fluidity, and linguistic competencies beyond our own experiences”, we need to urgently reassess our own hierarchical categorizations of historical textual phenomena and to study the linguistic environment of medieval England across the boundaries of individual languages:

The scribal class of medieval England, responsible in large measure for the enrichment of later Middle English, was in varying degrees a trilingual one. All three languages of Medieval England need to be studied as making up a unitary linguistic situation, especially in the later period when all three are often found in one sentence, rather than approached individually without reference from one to the other.  
(Rothwell 1998: 165)

### 6.1.2 The role of Latin in medieval England

Latin, the *lingua franca* of Christian Europe, remained the language of the Church and the ecclesiastical courts; to be officially literate (*litteratus*) meant knowing and using Latin. (Treharne 2011: 220)

In the twelfth century, most of the thousands of manuscripts composed and copied in England were produced in Latin, far outstripping the production of manuscripts in English or French (Treharne 2011: 220). While being the established language of literacy, Latin was also strictly a learned language, acquired through formal education and used as a spoken language only in ecclesiastical and academic contexts; since it was nobody’s first language, “hardly anyone could be expected to understand it spoken” in a public setting (Britnell 2009: 88).<sup>2</sup>

However, its highly standardized vocabulary, spelling and grammar made Latin especially well suited for use as a language of record even outside the ecclesiastical sphere, both in official institutions such as law courts and in financial accounts.<sup>3</sup> Even after the official spoken language of the English courts of law was changed from French to English in 1362, “legal documents were very generally issued in

<sup>2</sup> However, the Latinate clerks of the secular administration seem to have preferred Latin even for their own informal and utilitarian notes and memoranda written into the borough registers for their own reference (Britnell 2009: 82), which would seem to indicate a natural facility with the language at least in the context of administrative business. This should not be surprising, considering that the teaching of reading and writing was based on Latin and not the vernacular, creating a strong link between literate habits and the Latin language (Orme 1989: 170).

<sup>3</sup> Latin was used for the surviving medieval records of manorial and hundred court rolls (Brand 2000: 68), the borough court rolls (Britnell 2009: 81-2), and for the records of the city courts of London and Oxford, and most likely of other cities as well, from the end of the 13<sup>th</sup> century (Brand 2000: 68), most likely because they involved “definite formal requirements” which clerks were trained to fulfil in Latin, the “language the clerks were primarily expert in” (Britnell 2009: 81-2).

Latin” (Jefferson 2000: 185), and it seems to have been only in the 15<sup>th</sup> century that the secular royal bureaucracy and legal system began to employ English writing (Fisher 1977: 894).<sup>4</sup> However, at the end of the Middle Ages, it was only the church which continued to use Latin not only for its religious functions but also for its bureaucratic and legal writing (894), while in the secular sphere, Latin was rapidly losing ground as lawyers and government officials started to prefer French over Latin and creative writers began to favour English from roughly 1300 onwards (Clanchy 1993: 234).

### 6.1.3 Role of French in medieval England

While French was a spoken as well as a written language in medieval England, it never became the national language of England, and the loss of the French territories by 1204 and the multiple wars with France—culminating in the Hundred Years’ War (1337–1453)—were paralleled by the diminishing influence of French in both institutional and literary contexts (24–6). However, the form of French brought over by the troops of William the Conqueror in 1066, most likely a mixture of the various regional dialects of France (Rothwell 1998: 150), did take root on the English soil and over the next four centuries that it coexisted with Middle English, both influenced the native language to a significant extent, and was influenced by it, becoming a particularly English variety of French, *Anglo-Norman* (Rothwell 1985: 47). In spite of the rivalries and wars between England and France, the insular and continental varieties of French remained in contact with each other throughout the Middle Ages and Anglo-Norman did remain a living language in the upper echelons of society, at least in the southern part of England, playing a major role in the literate section of society (Rothwell 2004: 313–4).

From the 13<sup>th</sup> to the 15<sup>th</sup> century, the sociolinguistic identity of continental French changed in two ways: first of all, French in general came to replace Latin in many prestige domains, and secondly, the northern French of Paris emerged as the prestige variety in contrast to various rural dialects, including Anglo-Norman (Machan 2009: 365). Simultaneously, the Anglo-Norman variety was accommodating itself to the demands of the changing social context through the creation of new lexis independently of continental French and the semantic repurposing of inherited French lexis (Rothwell 1993: 23–4; Schendl and Wright 2011: 19). This meant that the two varieties took on different roles in the English linguistic repertoire, the native variety being what people spoke and the continental variety what they aspired to (368). The French influence on the English language operated mainly through the mediation of Anglo-Norman with very little direct influence from the Continent, which meant that while any continental French word may potentially occur in Anglo-Norman or in English, every ‘French’ word in English is most likely to be an Anglo-Norman one (Rothwell 1998: 152–3; Möhren 2000: 158, 166). Although Anglo-Norman was viewed as a ‘debased’ version of ‘proper’ French (*francien*) even by contemporaries, its importance in all spheres of life except for imaginative literature increases through the 13<sup>th</sup> century and into the 14<sup>th</sup>:

Long before the century was out, Anglo-Norman, alongside Latin, was

<sup>4</sup> An exception to this were the Great Seal and the Exchequer, which continued to use Latin well into the sixteenth century (Fisher 1977: 877).

the language used for recording the proceedings of Parliament, the *Statutes of the Realm*, the regulations governing the administration of numerous cities and towns; the laws of England, from the Conqueror onwards, could be read in Anglo-French, as could Magna Carta; the teaching and recording of English law was carried out by English lawyers through the medium of French; letters to and from kings and nobles, bishops and abbots, abbesses and prioresses, merchants and officials of both central and local government—a vast correspondence within the borders of England and also to and fro across the Channel—were couched in French. (Rothwell 1985: 46)

This wide use of Anglo-Norman French is especially remarkable, considering that intermarriage and the generally English-speaking environment meant that even among the gentry, “Francophone monolingualism may not have survived more than three or so generations after the Conquest” (120-1) and any remaining native speakers seem to have “lost the perception of themselves as French” before the end of the 12<sup>th</sup> century (Burnley 2001: 18). Based on the research of Rothwell, English seems to have become the “true vernacular”, of the English aristocracy as early as 1150 or 1200, while French was already a taught language, “learnt with a degree of formality from parents, tutors and clergy who knew and used it” (Orme 1989: 10-1, 169; Clark 1992: 120-1). The existence and popularity of didactic works for learning French, indicates that while French “was diminishing as a birth language, it retained its social desirability” (Machan 2009: 367).<sup>5</sup> Also the linguistic development of English culinary recipes seems to support the continued currency of Anglo-Norman in the everyday life of the literate classes, as the earliest surviving English culinary recipe collections, dating from the 13<sup>th</sup> and 14<sup>th</sup> centuries, are in Anglo-Norman (see subsection 8.1.1), the shift to Middle English—with obvious French influences—occurring only in the later 14<sup>th</sup> century.<sup>6</sup>

While Fisher (1977: 873) has argued that French would not have been the spoken language of the court after 1300, being increasingly relegated to a written role by English, French culture still retained its “prestige and allure”, and facility in French “remained a prized or necessary accomplishment in elite and other circles into the fifteenth century” (Bennett 2009: 332). For example Coleman (1981: 19) considers the language of Richard II’s court to have been French at least in terms of the books that were read, and in the 15<sup>th</sup> century Henry VI is still reported to have been proficient in French—which is not surprising, considering his French wife—leading Meale (1989) to conclude “that bilingualism, at least with regard to the written word, was not uncommon among several social groupings throughout the fifteenth century” (207). As was already observed above, the ability to read French does not seem to have been limited to the upper classes; for example

<sup>5</sup> Challenging this traditional view, Ingham (2010b) has suggested—based on his analysis of Anglo-Norman French as a *contact variety* of continental French and its influence on Middle English—that Anglo-Norman French was still spoken in the 13<sup>th</sup> and 14<sup>th</sup> centuries by “balanced bilinguals” of high social prestige, whose French was influenced by English, but was “deep-rooted and instinctive” (22) enough to also influence their English, which would have then—as the prestige variant—been imitated also by monolingual English speakers, resulting in the wider propagation of French influence.

<sup>6</sup> As Rothwell (1993: 44) points out, the fact that these recipes are clearly not translations from continental French but represent an independent tradition (Hieatt and Jones 1986: 860) testifies to the independence and productivity of the Anglo-Norman language long into the 14<sup>th</sup> century.

John Clerk, a 15<sup>th</sup>-century grocer and apothecary owned MS Harley 273, a composite volume collected over the 13<sup>th</sup> and 14<sup>th</sup> centuries that contained a variety of French religious, practical and literary works (Meale 1989: 207).<sup>7</sup>

One significant arena in which French was used almost exclusively throughout the Middle Ages was the secular legal system, which was “strenuously Francophonic” (Hanna 2011: 187). While formal records and documents were originally written in Latin, Anglo-Norman had most likely been the language spoken in the English law courts of all levels from the very beginning of the system of central law courts established by Henry II (Brand 2000: 66).<sup>8</sup> In the 14<sup>th</sup> century Anglo-Norman challenged Latin even as the language of law reports which recorded legal proceedings and were also used to teach the following generation of law students (Brand 2000: 75). The triumph of French over Latin for legislative purposes was not surprising, considering that “French was the working language of judges and lawyers and it made a great deal of sense to enact legislation in their professional language” (73).

The need to use French was not limited to lawyers and court officials, as Anglo-Norman was almost everywhere used as the language of petitions, not only in the parliament and the royal courts, but also for example when burgesses petitioned the mayor and the council of the city (Britnell 2009: 86). In the legal system, it is only at the level of the manorial courts that we find any evidence of the use of English in the proceedings of the 13<sup>th</sup> and 14<sup>th</sup> centuries (Brand 2000: 69), while the common law courts—despite the parliamentary statute of 1362 to the contrary—continued to plead in French until 1731. The only exception to this was the Royal Chancery, whose court conducted most of its proceedings in English from its very inception c. 1394 (Fisher 1977: 879-80), which had significant consequences to the English language as a whole, as will be pointed out in section 6.2 below.

In the 13<sup>th</sup> and 14<sup>th</sup> centuries French was also used as the language of record for administration, which meant that it “was not only spoken on a daily basis by large numbers of English citizens carrying out their professional duties”, but also “written in great quantity to keep the records needed by any advanced society” (Rothwell 1993: 21-2). This meant that when English took over these roles in the later 14<sup>th</sup> century, it retained much of the necessary vocabulary used by its predecessor, resulting in heavy lexical borrowing in the domain of administration (Rothwell 1985: 50).<sup>9</sup> In London, the use of French in administration was not lim-

<sup>7</sup> Another class of people who would be likely to have at least a working knowledge of French are soldiers and administrators who participated in the Hundred Years’ War, like John Fastolf, who managed to not only significantly raise his social standing, but also become ‘frenchified’ to the degree that they could fluently participate in French literate culture and bring back with them French books upon returning home (Beadle 2008). This would have resulted in many gentry households having strong linguistic, literary and cultural ties to France in the fifteenth century.

<sup>8</sup> Brand (2000: 66) sees this view—which is contrary to the older view which postulated a shift from English to French at some point—to be supported by the fact that no new technical legal terms were derived from English, and that French would have been the first language of the men appointed as the first royal justices as well as most of the litigants in the 12<sup>th</sup> century.

<sup>9</sup> Based on documentary evidence, Britnell (2009) has concluded that French was used even in late medieval urban administration for specific purposes, as 1) an oral language for proclamations and ordinances, 2) an oral language for taking oaths, 3) a language of recognized high status for communicating with external authorities, 4) a professional language for legal texts and the codification of custom; and among townspeople in general as 5) a polite language for devising written petitions presented to urban authorities, and 6) an “administrative language of ostentatiously high status out-



ited to governmental administration, as French was also used for the minutes and account books of many London craft companies long into the 15<sup>th</sup> century, until it was gradually replaced by Latin and English.<sup>10</sup> In addition to the administration of craft companies, also the vocabulary of many crafts themselves was heavily influenced by French; Jefferson (2000: 185) mentions goldsmithry as an example of a craft whose vocabulary at the end of the 14<sup>th</sup> century was almost entirely French, and the same would also seem to apply at least to some degree to culinary language, as is also apparent in the present edition, especially in the titles of the recipes.

Generally, the use of French in England—outside the legal profession, where it held its ground until the 17<sup>th</sup> century—seems to have peaked around 1350–1415, after which it was quite rapidly replaced by English, as “long-standing cultural conventions submitted to the logic of bureaucracy and practical convenience” (Britnell 2009: 89). Despite the decline in its official uses, French maintained its social appeal well beyond the Middle Ages, as the publication of numerous printed manuals for teaching French to English-speakers shows: 52 such manuals were published in the late 15<sup>th</sup> century and 139 in the 16<sup>th</sup> (Machan 2009: 369).

## 6.2 Vernacularization and standardization

While English had lost its status as the language of record for official purposes and for learned discourse with the Norman Conquest, this did not mean that English was relegated entirely to the spoken medium. It was largely the English monastic system that harbored a strong vernacular religious tradition throughout the twelfth century (Treharne 2011: 221).<sup>11</sup> However, the status of English as a literary language did suffer enough that it was not seen as an object of linguistic inquiry in the Middle Ages; the diversity of Middle English dialects was often acknowledged anecdotally, but no attempts were made to write either descriptive or prescriptive grammars, and the very idea of “linguistic correctness and fixity” was something associated exclusively with Latin (Machan 1994: 148). Even in comparison with other European vernaculars, “English in particular was characterized by lateness in the development of formal grammar and rhetoric”, and Machan (1994: 149–50) has argued that even when it came to be used in a variety of new roles

---

side the principal institutions of government” (87). Of these, he sees uses 3, 4 and 6 as examples of using the French language as a status symbol, while 5 implies that French was seen as an appropriately respectful medium for addressing mayors and councillors. Uses 1 and 2 he finds more difficult to account for, “given the practicalities of borough administration”, although he does suggest that the use of French for ordinances and oaths “may have had a certain symbolic or ritual value” (87).

<sup>10</sup> Being the language of international trade (Rothwell 2004: 318), it is not surprising that French was used by the company of Grocers as their normal language of routine records until 1428, by the Drapers until 1436, by the Merchant Tailors until 1445 and by the Mercers until 1459 (Britnell 2009: 87). Britnell sees this use of French by these mercantile associations as being “out of line with the normal practices of either manorial or urban recording” and conjectures that it might be “a status matter in the context of London society—a claim to be identified with aristocratic and professional legal coteries” (87).

<sup>11</sup> Current research indicates that the 12<sup>th</sup>-century tradition of English writing seems to have emanated almost entirely from the monastic institutions and the monastic cathedrals of the pre-conquest Benedictine Reform group (Worcester; Christ Church and St. Augustine, Canterbury; Winchester and Peterborough), produced most likely either for a non-Latinate in-house audience or for the pastoral use of parish priests affiliated with the institutions (Treharne 2011: 222).

during the 14<sup>th</sup> and 15<sup>th</sup> centuries, it remained essentially a spoken language even in its written forms (Machan 1994: 153).

### 6.2.1 Vernacularization of institutions

While Anglo-Norman was still “the language of the court, of chivalry, entertainment and legal transaction” in the late 14<sup>th</sup> century, it was already losing ground to English in noble circles, as indicated by the fact that Gower and Chaucer—commonly believed to have been writing for a noble audience—already considered it possible to write in English (Coleman 1981: 20).<sup>12</sup> Discussing the relationship between English and French in late medieval England, Bennett (2009: 333) suggests that “[a]t the royal court and aristocratic circles the balance between the vernaculars certainly tilted decisively towards English in the 1360s”, and Fisher (1977: 894-5) has found evidence that Richard II already preferred to read in English instead of French.<sup>13</sup> Thus, while French continued as the spoken language of some members of the nobility and merchant class into the 14<sup>th</sup> century, English had, by the time of Chaucer and Gower, become the domestic language for all social classes (Fisher 1977: 878-9).

The English language was also an intensely political issue in the 15<sup>th</sup> century, both in terms of the religious reforms of the Lollards and the last period of the Hundred Years’ War. The Lollards’ translation of the Bible into English was seen by the Church to pose a threat to the stability of both its institution and its ideology.<sup>14</sup> The threat of French overtaking English and destroying it if the French won the Hundred Years’ War was used as a political rallying cry in parliamentary addresses already in the 14<sup>th</sup> century (Machan 2009: 366-7), and when Henry V launched his second invasion of France in 1417, he changed the language of his correspondence from French and Latin to English. This decision, which Richardson (1980: 739-40) sees as a calculated move to win support for the war—was well-received as a patriotic gesture by the English-speaking middle classes who were funding the war and had little love for foreigners.

The emerging national identity of England in opposition to France and the development of national culture led to the recognition of English as a valid language of official business, codified not only in Henry’s change of language but also in various institutional decrees such as the 1362 Statute of Pleading, “which dictated that all court pleading should henceforth be conducted in English” and the 1422 decision of the London Brewer’s Guild to keep all of its subsequent records in

<sup>12</sup> Taylor (1996) has argued that while Margaret Beaufort (1443-1509), the mother of Henry VII, still had “diverse books in French” that she used for her devotional meditation, “by 1400 most of the English of all classes preferred books in their native language, and clerks across the land were busy producing them”, resulting in “a mass of late Middle English religious verse and prose, [...] most of it preserved in plain utilitarian volumes” (48).

<sup>13</sup> Rothwell (1985: 49) goes even further in accepting the view of M. Richter—presented in a paper, titled “Towards a Methodology of Historical Sociolinguistics” read at the World Congress of Historical Linguistics at Poznan in August 1983—that the English nobility had English as their mother tongue already by the latter part of the 12<sup>th</sup> century.

<sup>14</sup> Archbishop Arundel’s Constitutions from 1407 thus expressly forbade the translation of the Bible into English as well as the possession of vernacular translations dating to Wycliff’s lifetime, and a few decades later, even knowledge of the creed or prayers like *Pater Noster* or *Ave Maria* in English would be considered evidence of heresy (Machan 1994: 151).

English (Machan 1994: 150).<sup>15</sup> From 1362 onwards, English was also used for parliamentary addresses, as acknowledged by the parliamentary scribes (Fisher 1977: 879), the first English entry in the parliamentary rolls being a petition from 1388. English entries slowly proliferated over the next three decades, until they reach the majority from 1422 onwards and become the rule around 1450 (Fisher 1977: 880). A similar picture is painted by “any of the large classes of documents at the Public Record Office” that clearly shows a sudden change around 1430–35 “from a mere trickle of English documents among thousands in Latin and French, to a spate of English documents” (Samuels 1989: 80).<sup>16</sup>

In addition to administrative writing, this same process of vernacularization is also visible in learned scientific writing, as knowledge from several fields of scholastic learning—e.g. theology, philosophy, astrology, and medicine—was beginning to be disseminated outside the university world in English over the 14<sup>th</sup> century, although Latin naturally persisted along it in more learned texts well into the 17<sup>th</sup> century (Voigts 1984 and 1989; Pahta 2004b). The language of science used in these texts formed a new vernacular register at the end of the 14<sup>th</sup> century, widening the functions of English to the area of learning and adapting the Greco-Roman conventions of writing science to English (Taavitsainen 2000: 131; Pahta and Taavitsainen 2004: 1). Many Middle English medical texts were translations or adaptations from mainly Latin sources—such as Trevisa’s translations of Ranulf Higden’s *Polychronicon* and Bartholomeus Anglicus’ *De Proprietatibus Rerum*—but there were also original compositions in English, especially in the field of surgery (Voigts 1989: 381).<sup>17</sup> The first medical texts written in English were collections of medical recipes—which closely parallel culinary recipes in their textual conventions—but the scope soon widened to encompass general health guides, surgical texts and learned academic treatises (Taavitsainen 2000: 133). As Taavitsainen (2000: 132) has observed, this vernacularization of scientific writing seems to have been a part of the same conscious Lancastrian nationalistic policy as the shift to English in administration and the promotion of English literature that resulted in the “blossoming of English poetry and prose” (Coleman 1981: 24) in the 14<sup>th</sup> century.

As in the vernacularization of scientific writing, translation also played an important role in the production of vernacular literature, and for example many of the popular verse romances that were being produced as entertainment for the expanding literate class of the 15<sup>th</sup> century were translated from French. Although knowledge of French was apparently quite widespread among the literate classes, English was at the same time coming into its own as a literary language, and the decision to translate the works might have been based as much on a desire to develop the English language as to widen the readership of the texts (Meale

<sup>15</sup> The brewers reported that they decided to keep their records in English “because by the King’s use of it in his letters missive (that is, his informal correspondence) and in his other personal business ‘their mother tongue was beginning to gain lustre’ and because there were many of their craft who read and wrote English but did not understand Latin or French” (Machan 1994: 150).

<sup>16</sup> Also in the Year Book, containing reports of the pleas heard before the Common Bench, English began to intrude on the official French already in the reign of Richard II, creating a mixture of “French-language clichés and technicalities” and “English thinking” (Hector 1966: 25).

<sup>17</sup> Based on the preface of Trevisa’s translation of the *Polychronicon*, Trevisa and his patron Thomas Berkeley had “a fully fledged language policy with an ambitious goal, comparable to that of King Alfred”, which motivated the translation of these works (Taavitsainen 2000: 132).

1989: 217; Blake 1992a: 4). These changes—which were not unique to England but a part of a pan-European nationalistic movement (Taavitsainen 2000: 133)—reflected changes not only in the perception of the English language and its status, but also in social structure and social ideals, namely “a broadening of the middle range of society, its greater participation in government and its increasing demand for a literature read for information, for pleasure and for spiritual edification” (Coleman 1981: 24).

Although sociolinguistic accounts of late medieval England have often focused on the ‘triumph’ of English over Latin and French, the late medieval and early modern history of French does not seem indicate “anything like an agonistic struggle with English” (Machan 2009: 363). A more accurate description, based on the above mentioned evidence, could be that French simply fell out of use because the weight of tradition that had supported it so far could no longer compete against the fact that the majority of the population, up to its highest echelons had already been speaking English for quite some time.

### 6.2.2 Development of a standard

While Old English had developed a standard form of the English language, and a new standard would emerge again in the Early Modern period, the Middle English form of the English language “may be characterized as one where there was no generally agreed mode of writing it” (Lucas 1998: 170). The development of the Early Modern form of Standard English, however, is generally seen to have started already in the late Middle Ages. As Hope (2000) and others have pointed out, historians of English have until recently tended to consider the process by which this incipient standard developed and spread to be relatively well-established:

The standard, *as any fule kno*, is a non-regional, multifunctional, written variety, historically based on the educated English used within a triangle drawn with its apexes at London, Cambridge and Oxford. Even more specifically, the propagation of this ‘incipient’ standard can be linked to a particular branch of the late medieval bureaucracy: the court of Chancery. (Hope 2000: 49)

The basis for this established view can be traced back to the work of Eilert Ekwall in the 1950s,<sup>18</sup> which was subsequently elaborated by M. L. Samuels (see e.g. Samuels 1972 and 1989, originally published in 1963) and John Fisher (see e.g. Fisher 1977). In his 1963 article (reprinted in 1989), Samuels identified four different ‘incipient standards’ of written English that were “less obviously dialectal” (66) and could have served as starting points for standardization. The first of these, associated with Wycliffite manuscripts, was identified by Samuels as a literary standard based on Central Midland dialects. The three others, labeled as ‘Types II–IV’, were considered by Samuels to represent three consecutive stages in the development of London English and the precursors to Standard English.<sup>19</sup>

<sup>18</sup> Especially influential was Ekwall’s 1956 book, *Studies on the Population of Medieval London*.

<sup>19</sup> Type II was represented by eight 14<sup>th</sup>-century MSS—including the Auchinleck MS—that he localized to the Greater London area, Type III was the late 14<sup>th</sup>-century language of Chaucer and Hoccleve, which also closely resembles the London English recorded by Ekwall, and Type IV was the variety used by the clerks of the Royal Chancery (Samuels 1989: 67–71).

Based on his analysis of the three varieties associated with London using the ‘fit technique’ of dialectology used in the *LALME*, Samuels (Samuels 1972: 165-70 and 1989: 72-4) came to the conclusion that the dialect of London changed from a Southern one in the 13<sup>th</sup> century to an East Anglian one in the mid-14<sup>th</sup> century, and to a Central Midland one by the end of the 14<sup>th</sup> century. By associating these shifts to Ekwall’s earlier hypothesis about the immigration of East Midlanders to London and a postulating a second wave of influential merchant immigrants from the Central Midlands at the end of the 14<sup>th</sup> century, he established the idea of Standard English ultimately having developed from the dialect of Central Midlands. While the specific diachronic shifts observed by Samuels in London English and their connection to specific waves of immigration have been shown to involve some problematic assumptions (see e.g. Wright 1996), the broadly East Midlands character of London English, as well as the central role of London in the standardization of English have been generally accepted by historians of English.<sup>20</sup>

The most widely accepted hypothesis—repeated in recent textbooks on the history of English (see e.g. Baugh and Cable 2002: 192-5)—holds that the modern written standard of English emerged from the conventions established by the clerks of the Court of Chancery between 1420 and 1460—itself based on the form of English used by the Signet Office or private secretariat of Henry V (Richardson 1980: 738-9; Nevalainen and Tieken-Boon van Ostade 2006: 274)—and was propagated by professional scribes throughout England by 1460 (Fisher 1977: 896; Richardson 1980: 726). In terms of its dialectal characteristics, the variety used by Henry’s Signet Office and subsequently adopted by the Chancery was that spoken in the East Midlands. Summarising the reasons usually given for why this particular variant became selected as the standard, Nevalainen and Tieken-Boon van Ostade (2006) point out that it was not only spoken by a larger number of people than any other dialect, but was also used in an area that “was agriculturally rich”, “contained the seat of government and administration as well as the two universities Oxford and Cambridge”, “contained good ports”, and “was close to the archiepiscopal see” (275). These qualities made it a “high-prestige dialect in which the nation’s business is conducted” (Lass 1987: 61, quoted in Nevalainen and Tieken-Boon van Ostade 2006: 275), and thus ideally suited as a standard.

The establishment of a ‘standard’ Chancery English did not, however, occur overnight. For the decade after the death of Henry V (1422), the English-language documents of the Chancery “drifted linguistically toward what has been called Chancery Standard”, but still exhibited the kind of dialectal confusion expected of early 15<sup>th</sup>-century documents composed and copied by different scribes. It was only in the early 1430s that the Chancery had developed “a distinctive language, a coherent, standardized written dialect” which, with the prestige of the Chancery behind it, answered to the need for a standard dialect among lawyers, officials, legal scribes and litigious gentry, and slowly spread throughout the country during the middle of the 15<sup>th</sup> century, becoming the most commonly accepted written

<sup>20</sup> In addition to the methodological problems pointed out by Wright (1996), the attribution of the resemblance of late medieval London English to the Central or East Midlands dialects to demographic migration patterns can be seen to suffer from the assumption of causation on the basis of correlation; the fact that the dialect of London developed in the direction of the Midlands dialect does not need to mean that it was influenced by it let alone descended from it, but could be the result of entirely independent processes that merely produced a similar result for unrelated reasons.

dialect and the ancestor of the Early Modern standard (Richardson 1980: 726).<sup>21</sup>

While the London origins of Standard English have been generally accepted, the exact mechanism of its development is much more controversial, the above being merely the most popular hypothesis. While some scholars have followed Samuels in emphasising the role of the geographical patterns of migration to the capital, and others the fact that London was the national seat of government and justice and produced large amounts of bureaucratic texts, yet others have stressed the influence of the mercantile and civic elite of the city. For example Keene (2000), writing as a historian, sees the influence of London to lie not primarily in its status as “a site of government and power, but rather as an engine of communication and exchange which enabled ideas and information to be distributed and business to be done across an increasingly extensive, complex and varied field” (111). London was the place “where the greatest number of languages and language types, of both regional and overseas origin, were spoken and intermingled”, resulting in the emergence of one or more types of London English, although the “processes involved are far from clear” (Keene 2000: 94).

A further complication is added by the observation of Burnley (2001) that although a “distinct local language may be identifiable in a few documents, [...] most written records from the capital are dialectally diverse *mischsprache*”, and “it is in fact quite difficult to say what London English is beyond the combination of a few diagnostic features identified by scholars” (17). For this reason some scholars (see e.g. Wright 1996) have contested the view presented above, arguing that the ‘Chancery standard’ was in fact neither a cohesive standard, nor likely to have been the only or even principal precursor of the early modern Standard English (Wright 1996: 109). While Wright’s (1996: 108-9) critique of the methodology of Fisher (1977) makes it obvious that the traditional theory of Chancery English rests on seriously inadequate evidence, it does little to suggest an alternative to it apart from pointing out that the situation was probably more complicated than previously assumed.

Hope (2000) goes even further and challenges the whole ‘single ancestor-dialect’ hypothesis, which he sees as being based on a mistakenly biological, evolutionary tree view of language change. The two principal problems that Hope (2000) sees with this hypothesis are that first of all, the linguistic data does not seem to support it, as Standard English features “can be traced to an inconveniently wide range of dialects” (51), and secondly, languages and dialects are not like biological species that cannot interbreed with other species. While linguistic standardization is commonly defined as “the selection, elaboration and codification of a particular dialect” (Hope 2000: 51), Hope points out that process of selection is not in fact the selection of a single dialect, but of individual linguistic features from a range of dialects, which are then recombined and codified into a new dialect which does not have a single common ancestor. Instead of the ‘single ancestor-dialect’ hy-

<sup>21</sup> The diffusion of Chancery English is seen by Nevalainen and Tieken-Boon van Ostade (2006) to have involved two different mechanisms. First of all, the central role of the Chancery in both national administration and the legal system meant that a large number of literate people came into contact with Chancery documents. Since the Chancery represented “the only official body which attempted to produce a relatively uniform writing system”, its written forms came to be widely imitated by local administrators (276). Secondly, the Chancery offered “far more apprenticeships than jobs for clerks”, which meant that many trained apprentices returned home after their apprenticeship to find work in local administrative centres, taking their newly learned writing habits with them (277).

pothesis, he proposes an alternative one stating that “standardisation came about in English when changes in the medium (writing) and context (distance) of language use created a situation in which multiple instances of grammar competition occurred” (53). Because of the multiplicity of the processes involved, no single dialect provided all of the successful variants, but each feature involved its own process of selection.<sup>22</sup>

A similar argument has been made by Smith (2008), who has examined the relationship between the standardization of the morphology and orthography of English and the development of vernacular literacy and the changing role of written Middle English. According to him, the unstable and dialectally fragmented writing system of Middle English before the 15<sup>th</sup> century was largely the result of the predominantly local use of English, which made it “convenient to use spelling-systems which offered a fairly close grapheme–phoneme mapping” because they “eased the teaching of reading and writing by the ‘phonic’ method, which seems to have been usual for much of the Middle Ages”, resulting in the evolution of spelling systems “which reflected the wide range of phonological systems which existed in England during the medieval period” (Smith 2008: 215). When English gradually assumed also national functions—administrative, legal and commercial—at the end of the medieval period, the extensive written variation of the earlier Middle English period became inconvenient, resulting in a “communicatively driven process of dialectal muting” that “began to reduce the range of written variation” and resulted in a preference for ‘colourless’ usages employing spellings “in fairly wide currency” becoming dominant during the transition from Middle to Early Modern English (Smith 2008: 215).

As McIntosh, Samuels and Benskin (LALME: I, 47) point out, many dialects would seem to have had parallel, more ‘colourless’ forms for many dialectal words, indicating the parallel existence of two registers, one for local use and “an upper, more neutral register of the ‘colourless’ type for writings intended for a wider use or more exalted public” (135). The spread of a national standard spelling system can thus be seen to have involved not the imposition of a ‘foreign’ standard dialect across the country, but rather the gradual abandonment of idiosyncratic local variants in favour of parallel forms which were equally native to the local dialect but also in common use over a wider area.<sup>23</sup>

These kinds of observations have led Wright (2000b) to describe Standard English as a *consensus dialect* that has emerged over a long period of time, combining features selected from a variety of authoritative texts, and to argue that “there is no single ancestor for Standard English, be it a single dialect, a single text type, a single place, or a single point in time” (5), and that “no single late Middle English

<sup>22</sup> Rissanen (2000) has also emphasized the fact that in addition to different regional varieties, also different *genres* and *text types* contributed differently to the standardization process of different linguistic features: for example official legal documents and statutes seem to have exerted a strong standardising influence on other genres in terms of spelling, but in terms of syntax and lexis, they “adopted forms from other genres, decontextualised and deregionalised them, and thus marked these forms as part of the standard” (121).

<sup>23</sup> The association of standardization with the adoption of English for wider national purposes is also supported by the observation that the same process is also discernible in the history of French, which similarly subsumed many of Latin’s domains, becoming the language of national administration, and in doing so developed a prestige variant in the form of Paris French that became the basis of a standard (Machan 2009: 366).

or early Early Modern authority will show all the features that end up in Standard English” (6). Considering the fact that the Midlands dialects were considered to be relatively ‘neutral’ and understandable both to Northern and Southern speakers already by contemporary commentators like Trevisa (Taavitsainen 2000: 135), it is not unlikely that the result of this kind of functional levelling would resemble it rather than some other, more marked dialect, making this view of standardization quite compatible with the observed dialectal characteristics of the emerging standard.

Thus, it would seem to be more appropriate to see language standardization—contrary to the earlier view of selection and establishment of a single dialect—as “a set of ‘natural’ linguistic processes (selections, self-censorships) which are started when language users encounter formal written texts, and become unconsciously sensitive to linguistic variation”, triggering natural processes of competition “which operate independently for each linguistic variable”, producing a new hybrid dialect (Hope 2000: 52). As Wright (2000b: 6) points out, standardization is not “a linear, unidirectional or ‘natural’ development, but a set of processes which occur in a set of social spaces, developing at different rates in different registers in different idiolects”, and as such, much more complex than has been thought by earlier scholarship.

The effects of this standardization process are also visible in the language of the six *PD* manuscripts. As the analysis of their dialectal features in chapter 12 demonstrates, dialectal features that could have been considered East-Anglian in earlier texts were by the 15<sup>th</sup> century occurring in the same texts alongside features typical to other regions, indicating a gradual process of standardization. The language forms of the six *PD* manuscripts would seem to exhibit features expected of relatively standardized 15<sup>th</sup>-century English, displaying considerable similarities, most of which reflect dialectal features typical of the East Midlands.<sup>24</sup> However, the existence of clear differences in the dialectal profiles of the six versions despite their relatively late date reveals that the standardization process was still far from complete, and a large proportion of the English lexis still exhibited several morphological and orthographical variants. However, the occurrence patterns of orthographic and morphological variants in the six *PD* versions seem to be indicative of equivalence and co-occurrence rather than mutually exclusive and dialectally distinctive alternation, reflecting rather the normalization of heterogeneous idiolects and the weakening of dialectal identity than linguistic homogenization. In keeping with the observations above, the scribes do not seem to have translated the orthography of the recipes into a single coherent standard, but rather to have been content with employing—even within a single recipe—a variety of relatively ‘neutral’ forms, possibly originating in different dialects but acceptable also in many other varieties, indicating the kind of functional levelling, referred to above.

<sup>24</sup> Although the six versions also share many features that the *Linguistic Atlas of Late Middle English* (*LALME*) considers typical to the West Midlands.



## 6.3 Textual consequences of multilingualism

Medieval mixed-language texts are typically not the product of incompetent or non-native authors or scribes, but rather reflect the—often very high—multilingual competence of the authors and scribes who wrote or copied them. They often mirror consciously or unconsciously used discourse strategies which express a range of functions, including the multilingual identity of authors, scribes and/or the readership/audience of these texts. (Schendl and Wright 2011: 20)

While the vernacularization and standardization processes described above were in the 15<sup>th</sup> century quickly transforming England from a multilingual and multidialectal society towards the relative linguistic uniformity of the early modern period, the fluid multilingualism of the high medieval English culture still had significant consequences for the textual culture of late medieval England. As Trotter (2000) has pointed out, multilingualism was an essential feature of medieval English textuality, many writers clearly being “at ease in two or more languages”, with the widespread use of documents in two or more languages and individual documents containing a mixture of languages apparently creating no obstacle for efficient communication (2). This situation seems to have been most common at the turn of the 12<sup>th</sup> and the 13<sup>th</sup> centuries, when documents show an increasing amount of “linguistic fluidity and permeability” between the three languages, indicating that the literate traditions of Latin, Anglo-French and English were no longer separated, and that literate classes were largely multilingual between the three languages (Treharne 2011: 223). This situation seems to have persisted until the 14<sup>th</sup> century, with speakers and writers freely supplementing the resources of one language with those of the others and in many cases, eventually integrating the foreign elements into the English communicative repertoire (Möhren 2000: 166).

### 6.3.1 Multilingual manuscripts

The multilinguality of the literate classes is exemplified by the fact that many medieval ‘miscellany’ manuscripts combine texts in Latin, French and/or English “in no apparent order” and with no visual or codicological discrimination between the three languages, indicating that their scribes and users were relatively indifferent as to the language in which a text was written (Schendl 2000: 78, 2001: 310; Treharne 2011: 230).<sup>25</sup> This kind of mixing of languages seems to have been especially common in scientific and medical materials of the 14<sup>th</sup> and 15<sup>th</sup> centuries (Pahta 2004b: 35), as is demonstrated by the presence of material in more than one

<sup>25</sup> A good example of this kind of throughgoing multilingualism is provided by British Library MS Harley 978, written around 1260 by five scribes, most likely to the religious community of Reading Abbey, which contains “musical, scientific, medical, romance, satirical, didactic and prognosticatory material” in Latin, French and English, compiled from a variety of sources and put together from booklets, some of which were copied at Reading Abbey and others bought elsewhere (Treharne 2011: 227).

language in more than half of the 178 English scientific and medical manuscripts from the period 1375–1500 examined by Voigts (1989: 380).

Durham University Library MS Cosin V.iii.11, a 15<sup>th</sup>-century miscellany manuscript containing one of the versions of the *Potage Dyvers* recipe collection edited here, is a late example of this kind of a multilingual manuscript collection: in addition to the Middle English recipe collection edited here, it contains several collections of medical recipes in English and in Latin, medical miscellanea in both English and Latin, a trilingual herbal (or *synonyma*) in Latin, French and English, orthographical and grammatical advice and a sample letter in Latin, a collection of medical and veterinary recipes in English with some Latin, a treatise for a hermit or an anchorite and a fragment of an epistle in Latin, and some short verses in both Latin and English (see appendix F for a detailed description of the contents of this manuscript).<sup>26</sup>

### 6.3.2 Code-switching

As indicated by the description of multilingual manuscripts above, the complex multilingual situation in medieval Britain has been seen as extremely conducive to *code-switching*, i.e. the change from one language (or variety) to another within one act of communication (Schendl 2000: 77). Although this phenomenon has been extensively studied in the modern spoken context, it has only recently begun to attract attention from historical linguists, focused on written manifestations of the phenomenon. Before the mid-1990s, medieval texts exhibiting *code-mixing* were viewed quite negatively and seen as examples ‘broken’ or faulty language, resulting from imperfect language competence or idle attempts at word play (Schendl 2000: 79; Schendl and Wright 2011: 16–8). However, when the phenomenon began to attract more attention, scholars like Wenzel (1994) and Wright (1996, 1999) observed that the mixing of languages in medieval manuscripts is not “the result of imperfect language competence” or a random idiosyncrasy, “but a widespread specific *mode of discourse* over much of the attested history of English” (92). It was also discovered that it “occurs in a variety of domains, text types and/or genres”, both formal and informal, and that different genres or text types seem to employ different syntactic switching patterns and strategies (Schendl 2000: 85; Schendl and Wright 2011: 23, 28).

For example Wenzel (1994) noted that macaronic sermons featuring frequent code-switching from Latin to English are in fact carefully wrought following the formal conventions of the scholastic sermon and employ the structural features typical to the genre “with great care and technical skill” (74). Similarly, Wright (2000a) has noted the mixing of two or more languages to be the norm in medieval English business writing such as accounts and inventories, and both Wright and Rothwell (2000) have argued that the use of mixed language for administrative documents and municipal records of the fourteenth century was a “recognised policy, not merely the haphazard product of scribal ignorance” (230). The fact that code-switching between French and English is found in a letter to King Henry IV also “seems to point to the basic social acceptability of this linguistic strategy”

<sup>26</sup> The other miscellany manuscript containing a version of the *Potage Dyvers*, British Library MS Additional 5467, is more monolingual, consisting mainly of Middle English texts (both translated and apparently original) with only a single short item in Latin.

(Schendl 2000: 81) and to a lack of clear demarcation between English and French at the turn of the 14<sup>th</sup> and 15<sup>th</sup> centuries. According to Schendl (2001: 311), types of texts particularly conducive to code-switching include 1) mixed or 'macaronic' poetry, 2) drama, 3) business accounts, 4) sermons, 5) legal texts, 6) medical texts, and 7) letters and diaries.<sup>27</sup>

While research on modern code-switching has mainly focused on 1) the distinction between code-switching and borrowing, 2) the syntactic patterns, sites and possible constraints of code-switching, and 3) the functional and pragmatic aspects of code switching (Schendl 2000: 86), historical research has mostly been limited to the second of these topics. This is most likely because the unstable and thoroughly multilingual context of late medieval England makes the first distinction extremely difficult to establish in most cases, and effectively tackling the third topic would require more descriptive information on the patterns of code-switching in a wide variety of genres and text types than we currently have, as well as more contextual information about the pragmatic purposes, audiences and uses of the texts themselves.<sup>28</sup>

While the proportion of medieval English texts exhibiting code-switching is difficult to evaluate in the absence of a comprehensive inventory of Middle English mixed-language texts, some estimates have nevertheless been provided by scholars working on specific text types. Wenzel (1994: 2) has found that in the corpus of little over 500 surviving English verse carols of the 14<sup>th</sup> and 15<sup>th</sup> centuries, 210 introduce at least some Latin into the otherwise English stanzas. Code-switching to Latin is especially prevalent in religious writing: Wenzel (1994) observes that there is "probably no religious or devotional text in Middle English prose that does not include some Latin words, phrases or sentences" (5), as it was "not only customary but evidently *de rigueur* to quote authorities in their original Latin form" (6). Another field in which code-switching seems to have been very prevalent is scientific writing. In Pahta's (2004b) analysis of the code-switching practices in the medical sections (Books IV and V) of Trevisa's English translation of *De Proprietatibus Rerum*, she notes that despite its stated aim of making the information available to a non-Latinate audience, it makes regular use of Latin (37).

While Middle English recipes also make use of other languages, mainly French and some Latin, their use seems to be much more limited than either in medical treatises, business or administrative writing, or sermons. In the context of medical recipes, Hunt (2000: 135) has observed that by the fifteenth century most recipe collections are monolingual in English, code-switching and language mixing being much less prevalent than in the fourteenth century. In the *Potage Dyvers* (PD) texts, (Anglo-Norman) French is used mainly in the names of dishes, both for technical

<sup>27</sup> Of these, the most attention has so far been paid to scientific and medical writing (Voigts 1996; Hunt 2000; Pahta 2003, 2004a,b; Meecham-Jones 2011), administrative and business writing (Wright 1992, 1995, 1999, 2000a, 2010, 2011; Trotter 2003, 2010, 2011; Ingham 2011; Schendl 2011), and religious writing, especially sermons (Fletcher 1994; Wenzel 1994; Iglesias-Rábade 1996; Horner 2006; Hal-mari and Regetz 2011; Pahta and Nurmi 2011; Schendl 2013).

<sup>28</sup> As Schendl and Wright (2011: 28-9) point out, the second issue is also the one where historical research can most benefit from modern code-switching theories, as most of the theories dealing with the syntactic patterns and constraints of code-switching at least claim to be language-agnostic. Modern theories relating to the functional and pragmatic aspects of code-switching are less useful, as they have been developed not only for a different medium, i.e. spoken language, but also for a completely different social and cultural context.

terms denoting types of dishes and names of ingredients and adjectives denoting methods of preparation, as well as grammatical words linking these together. The fact that these three categories can occur in any combination (e.g.  $N_{En} + PP_{Fr} + N_{Fr}$ ,  $N_{Fr} + PP_{En} + N_{En}$ ,  $N_{En} + J_{Fr}$ , or  $N_{Fr} + J_{Fr}$ ) and are frequently abbreviated means that it is often extremely difficult to determine their linguistic status.

The use of Latin in the *PD* collections is limited to paratextual formulae like the *incipit* and *explicit* of the collection and recipe numbers, which are given in an abbreviated Latin form in MS D (e.g. *Ca<sup>m</sup>. ij<sup>m</sup>.* for ‘Capitulum secundum’). In the case of French technical terms, the reason for their use would seem to be simply the lack of a corresponding English term, the French term often—but not always—being in the process of being borrowed into English.<sup>29</sup> The fact that the code-switching practices of recipes are similar to those of business accounts is not surprising, as both text types are based essentially on the listing of items, which Wright (1999: 115) sees as a textual feature especially conducive to code-switching.

Schendl (2000: 82) notes that some Subsidy Rolls and business accounts that exhibit extensive code mixing also use heavy abbreviation to neutralize the specific characters of the languages involved and to blur the switch sites and the morphological clues about the linguistic status of the words. While some commentators have in the past condemned this practice as “‘degenerate’ and linguistically inadequate”, Wright (2000a) argues that it actually constituted “a functional register” which had the advantage of facilitating “the accurate reading and comprehending of business documents such as accounts and inventories, regardless of one’s mother tongue or precise competence in Latin syntax and morphology” (151). The degree to which this strategy is consciously used in the *Potage Dyvers* texts is not clear; most of the manuscript versions do not seem to abbreviate code-switched words any more frequently than they abbreviate others.<sup>30</sup>

Schendl and Wright (2011: 22) point out that like their syntactic patterns, the pragmatic functions of code-switching are often specific to a certain text-type. For example, whereas macaronic verses usually employ code-switching for humorous or satirical effect, the equally macaronic sermons use them for “serious moral exhortation”, making its function in these two contexts almost diametrically opposed (Wenzel 1994: 11). Similarly, code-switching or -mixing can be interpreted either as a means of intentionally creating *contrast* between the two languages, as in macaronic poetry, or as an indication of the *lack* of language demarcation, the resources of the second language being used freely to supplement those of

<sup>29</sup> In many cases the lack of morphological clues makes it impossible to decide whether the word should be considered English or French.

<sup>30</sup> The exceptions to this are MSS D and H4016, which abbreviate foreign words—as defined in chapter 10—roughly twice as frequently as they do words that are considered English: 49/22 per cent for MS D, and 29/17 per cent for MS H4016. This difference between abbreviating English and non-English words does not seem to correlate with the overall level of abbreviation or the proportion of foreign words in the text—the overall rates of abbreviation for the two texts are quite close to the average rate for all six texts (19 per cent), while the proportion of foreign to English words is very high in MS D (2.5%) and very low in MS H4016 (0.4%)—and although it does seem to correlate with the proportion of Latin of all foreign words (relatively high in both MSS), it is not seen in MS H279 which has a similarly high proportion of Latin, making it unlikely that the difference results from any systematic tendency to abbreviate code-switched words.

the primary language. Switches to Latin in the *PD* recipes are highly formulaic—occurring mostly in rubrics and other paratextual elements—and do not seem to fit either of these categories, being most likely a conventional text-organising device. Code-switching between English and French, on the other hand, clearly belongs to the second category, as there are no indications that the scribes copying the recipes would have considered the words here classified as French to belong to a different language.<sup>31</sup> While code-switching is often seen as a way of filling in lexical gaps in the primary language, it can also be used for terms which the speaker or writer knows in both languages, either for stylistic effect or for disambiguation by repeating the same concept in both languages (Hunt 2000: 132), although this does not seem to be the case in the texts edited here.<sup>32</sup>

One potentially fruitful way of viewing the use of French lexis and Latin paratext in the recipe collections has been suggested by Wright (1999: 115) in the context of Middle English macaronic business writing. She argues that the function of macaronic business writing was essentially to serve as a non-regional variety, a kind of national (and to some degree also international) *lingua franca*, used to record transactions in a way understandable to users of different linguistic backgrounds.<sup>33</sup> The use of seemingly French words for dishes or types of dishes, processes or ingredients that are not encountered outside of the culinary field could be seen as an example of a international professional lexis that would be equally understandable to a French-speaking and an English-speaking cook or culinary aficionado, while the use of Latin for incipits, rubrics and chapter numbers could be seen as the utilization of a shared bibliographical code, familiar to any medieval European reader, whether they could properly understand Latin or not.

### 6.3.3 Linguistic borrowing

The multilingual textual culture—especially between English and French—and the prevalence of code-switching also played a significant role in “the process of widespread relexification of English in the ME period” which peaked in the last quarter of the 14<sup>th</sup> century (Burnley 2001: 18).<sup>34</sup> As Ingham (2010a) points out, the “boundaries between the languages in medieval England were somewhat porous”

<sup>31</sup> This seems to have been the case with code-switching between English and French in general, as French was apparently regarded by the literate classes not as a discrete language entirely separate from English, “but rather as a part of the common stock of linguistic material available for use in records either in the form of complete lexical items or as components that could be combined with English elements” (Rothwell 1998: 163).

<sup>32</sup> In the first case, code-switching by a fluent bilingual is essentially “no different from style-shifting to the monolingual” (132); for example the use of French phrases among English could have a sociolinguistic function as an indication of status (Burnley 2001: 25). In the second case, code-switching is often used for didactic purposes to introduce Latin technical lexis by accompanying it with a vernacular gloss (Marttila 2011: 155–56).

<sup>33</sup> Wright sees this to be supported by the observation that macaronic business writing seems to have been most prevalent in a period of wide dialectal variation and to have disappeared with the appearance of a standard form of English, which would have replaced it as an efficient medium of communication.

<sup>34</sup> While the influx of French lexis to Middle English was earlier attributed almost exclusively to borrowing from mainland French through literate people travelling between France and England, a large proportion of the ‘French’ loans have been shown to be original innovations of Anglo-Norman by Rothwell (1998), indicating Anglo-Norman to have been a productive language well into the 14<sup>th</sup> century.

(2), which meant that both lexis and syntactic structures passed from one language to another quite fluidly. The fact that people passed back and forth between English and Anglo-Norman French in their daily business led “to specific lexical transfers both in the field of technical and general vocabulary” (86), which can be seen to go beyond mere ‘borrowing’ as Rothwell (2004: 314) has pointed out. Also the shift from Anglo-Norman to English in official contexts resulted in a significant amount of lexical transfer; for example when French gradually gave way to English in the *Rotuli Parliamentorum* over the fourteenth and fifteenth centuries, “the syntax changed, but a great proportion of the lexis was simply taken over as English, with or even without minor changes in spelling” (Rothwell 1985: 45).<sup>35</sup>

In addition to administrative vocabulary, also the technical lexis of everyday upper class life and courteous society, including cooking and dining, saw significant amounts of borrowing during the same period.<sup>36</sup> For example the traditional vocabulary of estate management was French, which meant that noble boys would be familiar with at least some French and promoted the mixing of the languages (Orme 1989: 157). Similarly, Coleman (1981) notes that chivalric and military treatises, which were popular entertainment reading the late 14<sup>th</sup> century were mostly derived from French models and borrowed heavily from “a highly technical chivalric French” (42-3). Based on this observation, the heavy borrowing of technical French terminology (including recipe titles) in Middle English culinary recipes could be seen as an indication that many medieval English dishes—or at least the practices of making them—originated in France, or that the English tradition of cooking was at least influenced by the French tradition (see section 8.1, and subsections 8.3.1 and 9.3.2).

In addition to the Anglo-French and Latin—whose influence is often difficult to distinguish from each other—Middle English was also influenced by lexical borrowing from a large number of other languages—including Scandinavian, Dutch, German, Greek, Spanish, Italian, Arabic and even some Indian dialects—either directly or via some other language (most often Latin or French) (Crespo Garcia 2000: 28).<sup>37</sup> In the culinary field, the most significant lexical contributors after French seem to have been Italian and Arabic, although often operating through the intermediation of French. Italian loan-words are of course extremely difficult to distinguish from French ones, but one likely candidate is the name of the sweet-and-sour dish called ‘egerdouce’; while the name could very well be French, neither

<sup>35</sup> Rothwell has written a considerable amount on the influence of Anglo-Norman on Middle English lexis, detailing the ways in which Anglo-Norman words were absorbed by Middle English (see e.g. Rothwell 1998, 2000, 2004, 2007).

<sup>36</sup> While English borrowed significantly from French via Anglo-Norman, Anglo-Norman itself remained relatively conservative, most likely because those who continued to use it did so “in large part because it was not English” (Marvin 2004: 17), i.e. in order to set themselves apart. While it picked up English syntax and was apt to lose distinctions not native to English, like grammatical gender, it adopted relatively little English vocabulary.

<sup>37</sup> Some of these languages were in use as spoken languages in England; for example Keene (2000: 109) estimates that in the late Middle Ages, about 10 per cent of the roughly 50,000 Londoners were of foreign origin, most of them identified linguistically as Dutch, coming from the northern Netherlands and the lower Rhineland. However, quantitatively these loans are of an entirely different order as the massive lexical influx from Anglo-Norman and Latin. It should also be noted that some of the words attributed by Crespo Garcia (2000: 28) to these languages were in fact established in Latin already in the Classical period, and could be considered thoroughly integrated into Latin by the time they were borrowed into English.

the word *aigre-doux* nor the dish itself are attested in medieval French sources,<sup>38</sup> while similar recipes do occur in Italian recipe collections (Hieatt and Butler 1985: 185). Perhaps the best-known Arabic contribution to Middle English culinary vocabulary is the name of the popular dish called ‘mawmenny’ (spelled in various ways), which appears to be a derivation of the Arabic dish called *ma’mūniyya* (Rodinson 1962, 2001a; Hieatt 1998a: 137).<sup>39</sup>

The close interaction between English, French and Latin in the medieval English context means that it is in many cases virtually impossible to decide whether an originally Latin or Anglo-Norman word is to be considered as foreign or as already integrated into English, making it extremely difficult to distinguish between borrowing and code-switching (Hunt 2000: 131-2; Rothwell 2000: 219; Schendl 2000: 81-2; Burnley 2001: 26; Schendl and Wright 2011: 26-7; Meecham-Jones 2011: 257-8; Trotter 2011: 182). While earlier research tended to treat code-switching and borrowing as two separate phenomena, “there is now a growing tendency to see them as situated on a continuum” (Schendl and Wright 2011: 24), especially in a multilingual context like late medieval London where the distinction between French and English may have been as vague for the medieval Londoner as it is to the modern scholar (Burnley 2001: 28).<sup>40</sup> In the case of mixed Middle English and Anglo-French in late medieval England, Trotter (2011) questions the usefulness of the whole concept of code-switching, since it assumes the existence of two separate code-systems between which language users can make conscious choices, arguing that the late medieval English vernacular in fact constituted a single code with lexical material inherited both from English and French.

While the use of Latin in some of the *Potage Dyvers* versions can be seen to constitute code-switching, the use of Anglo-Norman or French terms in varying states of assimilation is more appropriately seen as an example of the kind of “language mixing” (Trotter 2011: 183) described above, which precludes code-switching as a phenomenon by eliminating the required boundary between separate languages. Since the linguistic integration of a lexical item is “continuous rather than discrete” (Schendl 2001: 307), there are always borderline cases where the judgement between languages, especially English and French is somewhat arbitrary. For this reason, the identification of foreign-language elements in the present edition is based on a relatively conservative set of formal criteria that define a relatively inclusive view of Middle English and a high threshold for code-switching (see *Treatment of foreign words* in subsection 1).

## 6.4 Medieval literacies

While the concept of literacy—despite its everyday definition of being able to read and write, usually in one’s mother tongue—is acknowledged to be a pluralistic

<sup>38</sup> According to the *OED*, the word first occurs in French in 1541.

<sup>39</sup> The overwhelming preponderance of French as the contributor of English culinary vocabulary is demonstrated by the fact that Austin (1888) automatically interpreted this word as a French loan, explaining it as “apparently derived from the Fr. *malmener*, the meat being teased small” (136).

<sup>40</sup> For example Schendl (2001) and Meecham-Jones (2011) have suggested that the assimilation process of foreign words is a gradual one involving an intermediate stage where words “might be considered to be both ‘borrowed’ and ‘code-switched’ simultaneously” (Meecham-Jones 2011: 258).

concept even today,<sup>41</sup> it was an even more fragmented concept in the late medieval English context. First of all, the *multilingual* environment of late medieval England and the medieval hierarchy of languages means that the medieval concept of literacy was usually not connected to one's mother tongue. Second, the connection between reading and writing was much less natural in the manuscript age where writing implements were not ubiquitous and writing was its own specialized craft. Third, the more limited spread and extent of literacy, together with the variety of different functions it served meant that instead of the single comprehensive literacy we are used to in the modern first-world context, there existed a range of more limited functional varieties of literacy. These factors mean that instead of a medieval literacy in the singular, the phenomenon can be more accurately conceptualized as a variety of more or less related and socially, functionally and contextually conditioned *literacy practices* (Clanchy 1993).

According to Cerquiglini (1999), the research of the written culture of the Middle Ages has been misled not only by "the estimates of widespread illiteracy in the Middle Ages and the dubious consequences drawn from this", but perhaps even more by "the simplistic manipulation of the opposition between oral and written" (16). Many social and literary historians of the mid-20<sup>th</sup> century, perhaps most famously McLuhan in the 1960s, and Eisenstein and Ong in the following decades (see e.g. McLuhan 1962, Eisenstein 1979 and Ong 1982) saw the Middle Ages, despite its manuscript culture, as a predominantly oral society and celebrated the printing press of Gutenberg as the invention that shifted medieval Europe into a primarily written mode of communication (Saenger 1982: 367-8). Adopting the 'great divide' theory of Ong (1982), which was based on the observations of Goody and Watt (1968) on literacy in Ancient Greece and stipulated that literate and oral societies possess a different set of cognitive skills, this view saw literacy as a technological innovation—largely independent of the surrounding social context—that restructured these cognitive processes and brought about changes in society (Jones 2000: 42).

Clanchy (1993, originally published in 1979) criticized this view of literacy and printing as agents of social change and argued that literacy itself—or the 'literate mentality'—is not best seen as a technological invention, but a socially conditioned phenomenon that emerges as a response to the changing practical demands of the surrounding society and not from any "abstract desire for education or literature" (Clanchy 1993: 19). For Clanchy, the social development that led to the emergence of literacy in medieval England was the growth of administrative bureaucracy which could no longer be sustained through the use of memory but required written records for its functioning. However, as Jones (2000: 44) argues, the development of a literate mentality can happen at different paces and from different stimuli in different domains of culture.<sup>42</sup>

<sup>41</sup> United Nations Educational, Scientific and Cultural Organization defines literacy as the "ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts" and sees it basically as a social skill, involving "a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wider society" (UNESCO Education Sector 2004: 13).

<sup>42</sup> Following Clanchy's ideas, various scholars of literacy have since criticized the earlier 'autonomous' view of literacy and called for a more ideologically and socially contextualized approach (see e.g. McKitterick 1990; Street 1995; and Lowe 1998).



Instead of the traditional oral–literate dichotomy, this discussion of medieval literacy will adopt a view similar to that described by Jones (2000), based on the ‘ecological’ approach of Barton (1994) and on the “social theory” of literacy developed in Barton and Hamilton (2000).<sup>43</sup> This view, like Clanchy (1993) above, sees literacy not as a single, monolithic skill which an individual possesses to a greater or lesser extent, but rather as a selection of *literacy practices*—defined as “social practices associated with the written word” (Barton 1994: 37) or “the patterns, strategies and background information which individuals bring to literacy events” (Jones 2000: 51)—specific to a given society. These literacy practices are what allows a literate individual to participate in a variety of *literacy events*—or “communicative situations where literacy has an integral role” (Heath 1984: 71, quoted in Jones 2000: 50)—to various degrees and in various roles. As Barton and Hamilton (2000: 8–9) point out, this notion of *event* stresses the situated nature of literacy.

Jones (2000: 50–1) sees the benefit of this definition to be that it extends participation in literacy events also to individuals who would not be considered literate in the traditional sense, and as such is well suited to the description of the medieval situation where literate and oral practices interacted in intimate and complex ways. From the point of view of the present thesis, this view of literacy is also useful because it focuses on “the texts of everyday life” (Barton and Hamilton 2000: 9). For example culinary recipes are embedded in literacy events that mix both written and spoken language in the context of “a broader set of domestic social practices” with the practical and non-textual ultimate aim of preparing food (Barton and Hamilton 2000: 12). While medieval culinary recipes are decidedly written in their surviving form, they were intimately connected not only with the physical practice of cooking, but also with an oral tradition passed on from master to apprentice in the context of a medieval ‘professional’ kitchen or from mother to daughter in a more domestic context (see section 7.2), and could at various points of their transmission be communicated orally in the kitchen, written down by a household clerk, read aloud in the kitchen as a reminder, and studied in the library by the master or steward of the household.

Another fundamental difference between the modern and medieval conceptions of ‘literacy’ involves their different *foci* in terms of ability: whereas medieval assessments of literacy concentrate on the case of maximum ability, the literary and intellectual skills of the most learned scholars, modern ones have been mainly focused on the minimal ability of the masses to sign their own names or read simple documents (Clanchy 1993: 232). Many aspects of the medieval conception of literacy are “quite antithetical to any modern notions of what constitutes ‘literate behavior’”, especially to our idea that literacy requires “a command of multiple skills primarily ocular in nature”, since in the Middle Ages, reading was seen as a predominantly *intellectual* activity (Hanna 2011: 173):

<sup>43</sup> Barton and Hamilton (2000: 8) characterize literacy as a social practice through the following six propositions: 1) literacy is best understood as a set of social practices employed in events mediated by written texts; 2) there are different literacies associated with different domains of life; 3) literacy practices are shaped by social institutions and power relations, making some literacies more prominent and socially influential than others; 4) literacy practices are purposeful and involved in wider social goals and cultural practices; 5) literacy is historically situated; and 6) literacy practices change and new ones are born through informal processes of learning and sense making.

Discussions of medieval literacy have been bedevilled by the difficulty of distinguishing between the modern ‘literate’ and the medieval *litteratus*. When a knight is described as *litteratus* in a medieval source, his exceptional erudition is usually being referred to, not his capacity to read and write. (Clanchy 1993: 231)

The way in which literacy is evaluated and discussed in medieval sources may thus be one reason why historians have underestimated the extent of *minimal literacy* among the medieval peasantry and been “reluctant to allow such competence even to the gentry”, even though instances of smallholders, tenants and bondsmen authenticating written evidence by their signatures or seals are found already in the 13<sup>th</sup> century (Clanchy 1993: 233). Considering that the duties of male aristocracy—“to govern and defend the realm”—required them to understand administrative and legal documents (Orme 1989: 154), it seems certain that in the late Middle Ages, not only the male aristocracy but also the constantly expanding administrative class could read in the modern sense of the term, including them in the potential user group of practical texts like recipes.

#### 6.4.1 Defining medieval literacy

While modern literacy is associated with one’s mother tongue, the medieval term *litteratus* referred to the command of Latin, the learned language of the period (Hanna 2011: 173). While the original Roman meaning of the term had referred to a person with *scientia litterarum*, i.e. someone literate in the modern sense, it had by the 12<sup>th</sup> century become synonymous with the concept of *clericus*, its antithesis *illitteratus* being associated with the concept of *laicus* or ‘layman’,<sup>44</sup> as the only *litterati*—i.e. those with command of Latin—outside the Mediterranean area were Christian clergy (Clanchy 1993: 226-7). In the Middle Ages, both *clericus* and *litteratus* were relative and contextual terms and largely a matter of opinion, since they essentially meant ‘learned’ (229).<sup>45</sup>

While it is thus in the medieval sense an axiomatic truth that all laymen were considered illiterate, this does not mean that all non-churchmen of any particular place or time were unable to read or write (Clanchy 1993: 231-2). Neither were all churchmen necessarily able to read, write or speak Latin, especially on the lower levels of clergy. In the high Middle Ages, Thompson (1939) sees there to be “good reason to believe”, despite the “fragmentary evidence”, that the nobles of the Norman and Angevin England “were more familiar with Latin than is commonly supposed” (180). However, based on contemporary reports, there seems to have been a decline in the literacy of the noble laity from the end of the 12<sup>th</sup> century, connected to the rise of Anglo-Norman French as the literary language of the

<sup>44</sup> The *clericus/laicus* distinction is a Christian one between the ‘elect’ of God (Gk. κληρος, ‘kleros’, originally referring to a selection by lot) and the general mass of people (Gk. λαός, ‘laos’) (Clanchy 1993: 226-7).

<sup>45</sup> In the 14<sup>th</sup> century, the term *litteratus* thus had a multiple meaning; when applied to an ecclesiastic, it would have meant “that he was a university graduate with letters after his name”, and when applied to a layman, it would have indicated merely “that he was grounded in Latin grammar” (Coleman 1981: 24). This meant that a clergyman like a bishop could in some cases be referred to as a *laicus* or an *illitteratus* while a non-clerical member of the gentry could be a *clericus* and a *litteratus* (Clanchy 1993: 229).

nobility, although Clanchy (1993) believes most of the 13<sup>th</sup>-century gentry to have had some rudimentary teaching of Latin, allowing them to read simple Latin texts and thus be called ‘clergy’ (225).

Being overtaken by French as the literary language of high society, Latin seems to have become more an administrative language. Clanchy (1993) believes that by 1300, not only officials of the central government but also “manorial and village stewards, bailiffs, beadles, and reeves” (236) knew some Latin, as by that time the bureaucratic demands for Latin literacy had increased to the point where it was useful for any landowner to understand some administrative Latin (Clanchy 1993: 247). This means that the medieval equivalent of modern ‘minimal literacy’, i.e. being “able to read a little Latin, sufficient to get the gist of a royal writ or to understand a line in the Bible or in a chronicle”, would by 1300 have been “common among the gentry and may not have been rare among peasants” (Clanchy 1993: 246).<sup>46</sup> This situation seems to have persisted into the late medieval period, as Orme (1989) observes that literacy, in the sense of elementary knowledge of Latin, was “probably universal among the later medieval English aristocracy of both sexes”, based on “their involvement in keeping and using written records, in getting and sending letters, in owning books, and in a few cases even writing them” (170).<sup>47</sup>

### 6.4.2 Reading and writing

Another source of misconceptions regarding medieval literacy stems from the fact that in the Middle Ages, unlike today, the ability to read did not imply the ability to write, and the two were considered to be entirely separate skills (Clanchy 1993: 232). While reading and writing in the modern sense are commonly held to have a hierarchical relationship with reading as the primary skill, in the Middle Ages, the ability to write would not have meant that a person would be considered literate (271), and the ability to write was not considered an essential ability for a *litteratus*, who was defined by his ability “to read, understand, compose by dictation, make verse, and express oneself in the Latin language” (230). While reading was seen as an intellectual activity, writing was seen mostly as a craft, the two not being particularly intimately connected. Although Orme (1989: 73) is of the opinion that writing—either on wax tables or later on paper—was taught in grammar schools (see below), Hanna (2011: 173) argues that penmanship was a professionalized guild trade, only practised by those interested in composition and learned as a part of higher studies, and Britnell (1997) considers not only the records of the central administration but also manorial and urban business records to have been written by what he calls “professional scribes” (Britnell 1997: 16).<sup>48</sup>

<sup>46</sup> Although Clanchy (1993) has no firm evidence on the literacy of the peasantry before 1300, he finds the suggestion that some peasants were familiar with Latin “not implausible”, considering the role of the church in village life, which meant that at least theoretically “every adult in England should have known some Latin because of its use in the liturgy” (237).

<sup>47</sup> By the middle of the 15<sup>th</sup> century many London tradesmen were also being described as *litterati*, but as Clanchy (1993: 234) points out, this does not mean that they knew any more Latin than before, but rather that the criteria for literacy had simply changed.

<sup>48</sup> Fisher (1977) is also of the opinion that the skill of “writing and accounting in the fifteenth century was confined to a small, highly professional group” (896) and sees the earlier estimation of Kingsford (1962) that “the wives and sisters of country gentlemen could often write as well as their husbands

However, the multi-level authorial model described in subsection 2.3.1 means that the production of medieval manuscripts—whether administrative documents or culinary recipes—often involved not only the scribe serving as the *animator* of the document and possibly even the *author* of the text in the case of formal documents, but also the *principal*, whose interests were represented by the document. While it may be unlikely that the majority of household staff—or even the master of a household—would be able to write in the physical sense, they could still produce and make use of written documents for a variety of purposes. This means that we cannot restrict the group of people that produced, much less used, written documents to the category of professional scribes, but must expand it to cover also a wide variety of householders, craftsmen and businessmen who never themselves put pen to paper, but could nevertheless produce written material in the course of carrying out their business.

### 6.4.3 Education and acquisition of literacy

In 15<sup>th</sup>-century England, both children and adults who wished to acquire literacy had several options open to them, depending on their social class. Should they be fortunate to be born into a literate household, they could be taught the basics of letters, either by their mother (Orme 1989: 1), or in a wealthier family, a household clerk, mistress or schoolmaster (Orme 1989: 161). Even children of unlettered families could learn to read either as a part of their apprenticeship in the households of merchants, craftsmen, and shopkeepers (Orme 1989: 2), or by attending either a ‘petty school’ intended to provide an education for a merchant trade (Nevalainen and Tieken-Boon van Ostade 2006: 273) or a *grammar school* preparing one for a classical education and a clerical career, both of which were proliferating all around England in the early 15<sup>th</sup> century.

Aristocratic boys would also often be sent to the household of a greater magnate to serve as a page, which meant that they could be educated along the lord’s own children and wards by professional schoolmasters (Orme 1989: 6). The largest household involved in teaching boys was the royal household, which took in both common children “to help in the kitchens, storehouses and stables”, educating them to succeed their seniors and to replenish the supply of household staff, and noble children to serve as wards and pages, educating them as “courtiers, warriors and landlords able to give the king good service in adulthood” (Orme 1989: 2). In late medieval England, the court and household of the king thus served as an important centre of education and a repository of educated people. First of all, it served as a school not only for the royal children, but also for a host of other young aristocratic men and women, both orphan heirs in the wardship of the king and “noble children sent or gathered for the purpose” (Orme 1989: 153). Secondly, the court was a central meeting place for adult noblemen and -women who had been educated in other great households, schools and universities, which made it a thoroughly literate and even intellectual environment.

---

and brothers, and both they and their servants could and commonly did keep regular household accounts” (35) as overly optimistic.

### Elementary education<sup>49</sup>

From the time of the Norman Conquest, there seems to have existed a system of ‘town schools’, intended for “males, ranging from boys to young adults, who intended to become priests, monks, administrative clerks, or literate laymen” (Orme 1989: 4). During the 12<sup>th</sup> century, there evolved the specialized profession of the schoolmaster (*magister scholarum*), specialized in the teaching of literacy skills (49).<sup>50</sup> In late medieval England, Orme (1989: 50) evaluates the total number of schoolmasters to have still been quite low, setting the total somewhere between 300 and 600, which meant that the average English town had only one schoolmaster and London and Oxford perhaps a dozen.<sup>51</sup> Children could begin their tuition at a very early age, sometimes as young as three or four,<sup>52</sup> beginning with learning the alphabet and then proceeding “to study liturgical texts like the psalter and the antiphonal, learning how to spell the words, how to pronounce them, and how to sing them to the rules of ecclesiastical music” (Orme 1989: 170). This level of competence would already have guaranteed that they could pronounce—although not necessarily understand—the words of a Latin text and use a prayer book in church, and perhaps even more importantly, provided a basis for “understanding texts in the languages they spoke”, i.e. English and French (171).

In a *grammar school*, which could take the form of a free-standing grammar school, monastic almonry school, or a private tutor, the student would have proceeded to learn the rules of Latin grammar and some Latin literature (Coleman 1981: 26). In addition to reading, pupils may also have learned to write, although this may have referred to the act of *composing*, i.e. producing prose and verse text, rather than to actual penmanship.<sup>53</sup> While the language taught in grammar schools was Latin, the language of oral tuition—at least according to Ranulph Hidgen, writing in the 1320s—was Anglo-Norman French until about the middle of the 14<sup>th</sup> century, presumably as a legacy from the 12<sup>th</sup> century when schools had primarily catered to the French-speaking elite (Coleman 1981: 30; Orme 1989:

<sup>49</sup> Unfortunately, information about the medieval educational system is scarce, since the “history of English school education before the Reformation is largely a matter of local history”, requiring the history of English education to be pieced together from widely scattered local records (Orme 1989: 33).

<sup>50</sup> During the 13<sup>th</sup> century, at least 70 settlements in England are known to have had a school, including not only London which had three, but also all the cathedral cities, as well as “many country towns, ports and market towns” (Orme 1989: 5).

<sup>51</sup> In the later Middle Ages, these ‘town schools’ formed an important part of the educational establishment and resembled modern schools in the sense that they were “self-contained, open to the public, and taught by specialist teachers” (Orme 1989: 6).

<sup>52</sup> For example Philippa and Blanche, the daughters of Henry IV (then earl of Derby), were three and five, respectively, when copies of ABC were bought for them in 1397. Margaret Plumpton, was learning the psalter at the age of four in 1463, and Edward V had a schoolmaster when he was five in 1476 (Orme 1989: 170). However, literacy was not always acquired in childhood; for example William Smith, a Leicester Lollard in the 1380s learned to read and write as an adult and proceeded to even write books on religious topics (Orme 1989: 7).

<sup>53</sup> According to Orme (1989), “[t]here can be no doubt that writing was practised in schools from their earliest days”, and from at least the 12<sup>th</sup> century pupils of schools “practised how to write, composed written exercises in prose and verse, made notes and copied extracts from the standard texts” (73) using wooden tablets covered with wax. Unfortunately, wax tablets do not survive very well and it is difficult to establish the degree to which this was done; it is only from the fifteenth century that we begin to have surviving personal work-books of masters and pupils, written on the newly available and increasingly affordable paper.

10; Hanna 2011: 181).<sup>54</sup> However, in 1385, Trevisa reports that the language of instruction in grammar schools all over England had changed to English (Coleman 1981: 30; Hanna 2011: 181-2). Soon after 1400, English was being used both for Latin grammar books and instruction in grammar schools, affecting not just the schools but also the English language; by explaining Latin grammar in English, using English parallels, the teaching made the pupils think of English in Latin grammatical terms (Orme 1989: 11-2).

Until about 1400, attendance both at grammar schools and the university was limited to “aristocratic boys intended for ecclesiastical careers” (Orme 1989: 166), but in the late 14<sup>th</sup> century the hold of the clergy on education began to loosen as “lay benefactors, guilds and municipal bodies took part in founding and governing schools”, promoting the spread of lay literacy and resulting in more boys attending school without any intention of becoming priests (23).<sup>55</sup> When Henry VI came of age in the late 1430s, he also took an active interest in the development of grammar schools, followed by the nobles of his court, resulting in what could even be called an educational movement (14), leading to the founding of more than a dozen schools during the late 1430s and the 1440s.<sup>56</sup> This was a major development in the spread of lay literacy and meant not only that a wider than ever group of people had the chance to acquire literacy regardless of their economic means (26, 35), but also that the pupils of medieval schoolmasters were no longer just future clergy, but “ranged from those who wished merely to read and spell, to those who sought the fluency in Latin that was necessary to enter university” (63). This made it possible for theoretically anyone bright enough to acquire literacy and find themselves in significant positions within a major household (Coleman 1981: 24). In addition to the authorized town schools, private and informal tuition was also available in the larger towns, where independent masters instructed pupils in private houses. Those who could not gain access to schools, such as “women, people in remote areas and the poor [...] could learn some letters from a local priest or clerk or lay person” (6).<sup>57</sup>

### Higher education

Those aspiring for more advanced literary skills and/or for a clerical career continued their education in the university, the two options in England being Oxford and Cambridge. Students usually entered the university at 14–16 years of age, and their first step was the four-year arts course, which led to a bachelor’s degree and was considered to be the foundation of all learning (Hanna 2011: 183). Those who felt a further calling for an intellectual life might stay for a further three-

<sup>54</sup> This traditional use of Anglo-Norman in elementary education helps to explain the persistence of Anglo-Norman in upper-class circles, and also meant that English speakers were doubly challenged, needing to first learn French in order to benefit from grammar school education.

<sup>55</sup> By the early fifteenth century, boys “definitely intended to remain seculars” were also being sent not only to *grammar schools*, but also to the universities, including the eldest sons and heirs of nobility and gentry who sought university education to enhance “their adult careers as landowners and men of affairs” (166).

<sup>56</sup> This interest in elementary education also coincides with the first recorded examples of the eldest sons of peers being sent to study at Oxford and Cambridge, displaying an appreciation of formal higher education by the lay aristocracy (Orme 1989: 15).

<sup>57</sup> There are several records of priests with pupils from all over the country and parish clerks are mentioned teaching boys at various churches in the 15<sup>th</sup> century (6).

year course leading up to a master's degree, which conferred them the right to teach university students anywhere in Europe. A small minority went on—after a year or more as a member of the teaching staff—to one of the three 'professional schools', medicine, civil and canon law, and the highest of all, theology.<sup>58</sup> Most university students, regardless of whether they graduated or not, entered service either with the church or with a secular upper class household, which were always in need of educated men to serve as reeves, factors, and "bureaucrats of all sorts" (183). While the details of administration were not taught at the university, any university student would have picked up—to varying degrees—a set of 'transferable skills' that made him suitable for the position of an administrator (Hanna 2011: 189-90).

The universities were not, however, the only place to acquire more advanced literacy skills. Education in the kind of practical literacy required in royal or municipal government, including the ability to efficiently read and draft documents, was also provided by the administrative offices themselves. Before university education became popular among lay aristocracy in the 15<sup>th</sup> century—and even after that—a more important role in the education of gentry and aristocracy was thus played by the Inns of Court and Chancery (Britnell 1997: 7).<sup>59</sup> Originally intended as an institution for the training of professionals, in the early 15<sup>th</sup> century they became "infiltrated by nobility and gentry merely intent on acquiring a general education for aristocratic life" (Orme 1989: 167). Most of these aristocratic youths were the sons of knights and gentlemen, "for whom the study of law conferred real advantages in running the affairs of a small landed family" (168), and who used their legal education as the basis for the career of a 'man of affairs'.

#### 6.4.4 From aural to silent reading

In his important study of vernacular literary texts, Chaytor (1950) saw medieval communication to have been primarily oral, and maintained that it was the invention of printing that was chiefly responsible for the transition from aural reading to silent reading at the very end of the Middle Ages. As was pointed out above, this idea was taken up by McLuhan (1962) and his students and became part of the received wisdom about medieval reading, although even McLuhan himself admitted that "nobody has ever gathered adequate data on this question" (84).

While the book culture of Ancient Rome and Greece had been based on oral reading and dictation, the shift from reading aloud to silent reading seems to have begun in the monastic scriptoria in the 7<sup>th</sup> to 9<sup>th</sup> centuries (Saenger 1982: 373-4), developing into true silent reading with the eyes alone "with the evolution of a more rigorous intellectual life in the 12<sup>th</sup> and early 13<sup>th</sup> centuries in the *studia* of Cistercian abbeys and at the cathedral schools of the eleventh and twelfth centuries" (384).<sup>60</sup> As Chartier (1995) points out, the precondition of silent reading

<sup>58</sup> For a description of the course of medieval university studies and the teaching methods involved, see Hanna (2011: 183-7).

<sup>59</sup> For example the specialized literacy practices—mainly in Anglo-Norman French—required by the legal profession were taught in the Inns of Court and the Inns of Chancery, although the universities, Oxford in particular, also provided what could be seen as "associated schools of paralegal studies", which taught students French and skills such as accounting and legal drafting, as well as *ars dictaminis*, the ability to compose official correspondence (Hanna 2011: 187).

<sup>60</sup> Visual representations of silent reading are found already in the early Middle Ages, some of the ear-

“was the separation of words by Irish and Anglo-Saxon scribes during the high Middle Ages, and its consequences were considerable, creating the possibility of reading more quickly, and so reading more texts and more complex texts” (15-6). The transformation from an oral monastic culture into a written scholastic one over the 13<sup>th</sup> century “had at first only a limited effect on lay society, particularly in northern Europe” (Saenger 1982: 405), and it was only in the 14<sup>th</sup> century that silent reading spread from the universities to the world of lay aristocrats (Saenger 1982: 405; Chartier 1995: 15-6; Taylor 1996: 43). While private silent reading became more and more pervasive in the 14<sup>th</sup> and 15<sup>th</sup> centuries, public oral reading still continued to play an important role for example in university lectures (Saenger 1982: 391).<sup>61</sup>

Also writing—in its modern sense of composition—was before the 14<sup>th</sup> century “associated with dictating rather than manipulating a pen” (Clanchy 1993: 270-1).<sup>62</sup> While the physical act of writing was considered a separate skill from the composition of the text, they became more closely integrated in the 14<sup>th</sup> century as the silent composition of texts directly into writing became more common, also changing authorial expectations regarding their reception: “when texts were composed silently, authors expected them to be read silently” (Saenger 1982: 390-1). This meant that the changes in reading practices also became visible on the manuscript page:

The complex structure of the written page of a fourteenth-century scholastic text presupposed a reader who read only with his eyes, going swiftly from objection to response, from table of contents to the text, from diagram to the text, and from the text to the gloss and its corrections. (Saenger 1982: 393)

Because silent reading allowed for greater freedom in the movements of the eye, it favoured the perusal and *reference reading* of books (Saenger 1982: 385), which increased the popularity of encyclopaedias and other reference works that could be characterized as *discourse colonies* intended for nonsequential reading. This kind of “discontinuous and disjunctive” reading, well suited for encyclopaedias, herbals and recipe collections has been seen as typical of the Early Modern period (Knight 2009: 119), but it seems to have emerged already in the late Middle Ages.

However, as Cerquiglini (1999: 16-7) points out, this shift did not imply that oral communication diminished in quantity, but rather that it began to lose its authoritative character. The oral practices of reading aloud and dictating also survived because they had the benefit of permitting also “the non-literate to participate in the use of documents”, making the clerk or scribe effectively “a medium

---

liest occurring in conjunction with the Canons of Eusebius, which constituted “a primitive chapter concordance of the New Testament”, clearly intended for silent reference reading (Saenger 1982: 375).

<sup>61</sup> Although here as well, visual reading was indispensable, as students were expected to follow the lecturer by silently reading the text from their own books (Saenger 1982: 391).

<sup>62</sup> This is also reflected in the terminology used for it; whereas *scribere* and *dictare* had earlier both been used to refer to the act of composition, by the tenth century “*dictare* had clearly supplanted *scribere* as the standard synonym for composition”, *scribere* referring “almost exclusively to the physical act of writing” (Saenger 1982: 380). The art of literary composition through dictation was governed by the *ars dictaminis*, taught in schools as a part of rhetoric (Clanchy 1993: 271).



between the speaker or hearer and the document” (Clanchy 1993: 271). This made medieval literary culture “strongly ear-centered, auditory, and performative”, people being “accustomed (and expected) to engage” with texts orally, extending literate practices beyond the circle of those literate in the modern sense; people incapable of ‘eye-reading’ were often able to memorize, selectively recall and perform large amounts of text (Hanna 2011: 174). This has important implications for the production and use of utilitarian manuscript texts like culinary recipes: the surviving written documents are likely to represent a rich and multi-faceted tradition of communicative events involving a variety of people employing a variety of both oral and literate practices.

### 6.4.5 Changing functions of literacy

As the quantity of written material produced by the society increased significantly and writing came to be seen not just as “a means of conservation and memorization” but as a tool for intellectual work (Chartier 1995: 16), a new style of reading became the norm. It was less linear than before and based not on memorising text sequentially, but rather on conveniently locating a specific piece of information from a large mass of text. To answer the demands of this less linear style of *reference reading* in the 13<sup>th</sup> to 15<sup>th</sup> centuries, texts were being divided and subdivided into smaller sections, or *distinctiones*, below the traditional chapter division and provided with navigational aids such as tables of chapter headings and alphabetical subject indices (Saenger 1982: 392), also changing the ways in which texts were being constructed and encoded into physical documents.

The gradual spread of literacy down the social scale meant that written texts were used by a more varied group of people than before. This resulted not only in the blurring of division between the traditionally literate aristocracy and clergy and the newly literate commoners, and in “the elevation of merchants, financiers, and professionals” to “positions of great power and influence” (Amos 2001: 23-4), but also in changes in the functions served by literacy itself. Coupled with the decreasing cost and increasing availability of writing material after the introduction of paper into England in the 14<sup>th</sup> century and its proliferation in the 15<sup>th</sup> century (De Hamel 1992: 16; Britnell 1997: 18), this meant that written documents began to be used for a wider variety of purposes than ever before, which further increased the importance of literacy within society. In describing the diversification of the function of literacy in the late Middle English period, Coleman (1981: 25) distinguishes three different kinds of literacies, reflecting different levels of literate ability:

- 1) The *pragmatic reader* who could read and write only in the narrow context of his business transactions, his reading and writing mostly consisting of frequently recurring formulae which he had memorized and could apply, without necessarily knowing the meaning of individual elements of a sentence.
- 2) The *cultivated man or woman* who read for private entertainment and edification, representing either the nobility or the urban middle classes.
- 3) The *clerk*, who was a professional reader and writer and had made these skills into a career in the church, the civil service or the law courts, usually

after some studies in the university.

In addition to the level of individual ability, literacy practices have also been similarly categorized according to their function. Many scholars of literacy (see e.g. Britnell 1997 and Butcher 2004) have followed Parkes (1991, originally published in 1973) in distinguishing *practical* or *pragmatic* literacy, or the skills required to meet the needs of a growing volume of written business, from the more ‘literary’ literacy, or reading for general instruction, edification or entertainment. From this point of view, a *pragmatic text* has been defined as a context-bound document which “contributed to some legal and administrative operation and was produced for the use of a particular administrator or property-owner” (Britnell 1997: 3), while a *literary text* “had the capacity to instruct, edify or entertain an indefinite number of readers” (Britnell 1997: 3). However, since these definitions group together reading for entertainment and reading for the acquisition of practical knowledge, restricting pragmatic literacy to administrative and business texts, it is not very useful for understanding the literacy practices related to utilitarian texts like recipes and other household literature.

In this regard the original definition of Parkes (1991b) for pragmatic literacy as “the literacy of one who has to read or write in the course of transacting any business” (275) is actually more useful, as it can be seen to include also reading for the acquisition of practical information required for such business. In order to distinguish between reading for pleasure or moral edification and reading for practical profit—which can be motivated by very different impulses and thus potentially associated with different classes of people and social contexts—I will suggest the following three functional categories of literacy:

- *instrumental literacy*, covering the use of administrative records and business documents as well as legal writs and official proclamations in the process of performing specific economic, legal or social transactions, whether by preparing such documents as a record of an event or transaction, or using them for their evidentiary or record value;
- *utilitarian literacy*, covering the use of written documents for the acquisition of knowledge required for some practical task or operation or for learning new skills, whether it be the manufacture of a product or the successful completion of a process;<sup>63</sup> and
- *literary literacy*, covering the use of written documents for entertainment, edification or for acquiring or communicating abstract knowledge about the world.

While these types of literacy are defined on the basis of their function, their ‘passive’ use can also be seen to represent progressively increasing ability.<sup>64</sup> While instrumental literacy requires merely that the person understands the function of the written document and has a rough idea of its content, utilitarian literacy in-

<sup>63</sup> This category is considered to include also ‘practical’ *devotional literature* like Books of Hours which provided a framework for their daily cycles of private meditation and served as practical aids for preventing distraction and focusing one’s mind on the contemplation of the life and passion of Christ.

<sup>64</sup> ‘Use’ is here referred not to the production of a text but rather the use of an existing text for its intended purpose, i.e. arguing for the ownership of a parcel of land by using a land grant, obtaining a piece of practical information from a manual, or deriving pleasure or inspiration from a literary, religious or philosophical work.

volves the ability to decode the literal meaning of the document, and literary literacy requires the reader to not only decode the literal meaning but to understand and appreciate its allegorical significance. These three types of literacy are also differently motivated: instrumental literacy is motivated by the external demands of functioning in society, utilitarian literacy by a need or desire for practical benefits, and literary literacy by a desire for either enjoyment or moral, spiritual, or intellectual improvement. In addition to relating the literacy practices involved in the use of recipes and other utilitarian texts to the more established and better-studied practices of instrumental or pragmatic literacy and literary literacy, these categories will also be used in subsection 8.3.2 as a basis for a system of *genre types* used for contextualising recipes and other types of utilitarian literature as a *genre*.

#### 6.4.6 Development of vernacular lay literacy

The spread and diversification of literacy practices was not limited to the institutional and instrumental use of writing, as the 14<sup>th</sup> century saw an increase in the number of not only professional clerks, but also of those people who read for their own practical benefit, pleasure and edification, and literature which had previously been read only by the scholarly elite in Latin or French was being translated into English for the consumption of the rising middle class (Coleman 1981: 39–40). As several scholars (see e.g. Meale 1989 and Sponsler 2001) have pointed out, the 15<sup>th</sup> century saw a shift in the English middle classes from *pragmatic literacy* to a ‘non-pragmatic literacy’:

Those who, whether members of the gentry, or professional or mercantile classes, had previously acquired the skills of reading and writing in order to conduct their business affairs with greater efficiency, were now increasingly directing those same skills to other ends, using the written words for edification and entertainment. (Meale 1989: 217)

While this vernacular lay literacy is generally seen to have emerged in the late 14<sup>th</sup> century, its roots go back to the 12<sup>th</sup> and 13<sup>th</sup> centuries when the traditional division between cleric and lay, literate and illiterate, was broken down, resulting in literacy in the sense of minimal Latin becoming commonplace. Since literacy had been associated with Latin for a thousand years and also elementary teaching had been conducted in Latin, it was only after the spread of rudimentary Latin literacy had made the literate mentality familiar to a significant portion of the populace that “literacy could become a common vernacular habit” (Clanchy 1993: 251). Coleman (1981) sees this late 14<sup>th</sup>-century lay literacy to have consisted of “an ability to read and write in English and perhaps in either French or Latin” (24). According to him, reading English “appears to have become an assumed skill”, the term ‘unlettered’ being used to describe someone as unable to read Latin and possibly French, although the lack of direct documentary evidence about vernacular literacy makes its extent difficult to estimate.

However, the large number of English translations of both French romances and works discussing contemporary religious and political issues, as well as the popularity of the Middle English poetry of Chaucer, Gower and Hoccleve indicate a growing demand for Middle English literature. This demand was not limited to just entertaining literature, but extended also to utilitarian writing, such

as scientific and medical treatises, which were being translated in great numbers mainly from Latin.<sup>65</sup> While this influx of new vernacular texts—both literary and utilitarian—can partially be explained by the increasing preference for English among the traditionally literate upper classes, evidence from the wills of more middle-class burgesses and rural landowners shows them to have been literate at least in the vernacular and to have “displayed a zest for edifying English texts” (Coleman 1981: 23). Meale (1989: 217) also sees the existence of a wide variety of manuscript miscellanies—containing both vernacular and Latin texts and presumably copied for the personal use of late medieval householders—as an indication of the general increase of literacy.<sup>66</sup>

As a consequence of this expanding readership, “the production of books of all kinds came to be a profit-making activity, and books began to be marketed in ways that appealed to the readership with a vested interest in commercial activity and self-enhancement” (Sponsler 2001: 4). Literacy for recreation or self-improvement—which after 1300 meant mostly vernacular literacy—was also becoming more useful, “as more and more was being written down in vernacular languages” (Clanchy 1993: 247), including treatises on things like accountancy and estate management, which would be very useful for a nobleman or a gentleman. The social status of reading and book collecting, together with the manifest interest in works of practical nature and the existence of a relatively large literate upper middle class—both in the city and in the country—would have created a large and varied potential readership for utilitarian texts like culinary recipe collections.

<sup>65</sup> As Jones (2000: 36) points out, the demand for vernacular translations of learned medical texts indicates the emergence of a new kind of literate audience which was interested in learned medical knowledge but was not comfortable in receiving it in Latin, although she also notes that the shift towards the vernacular was not instigated exclusively by the non-Latinate, but was also advocated by the more educated, for whom the motivation for choosing vernacular texts was not their inability to read Latin Jones (2000: 36).

<sup>66</sup> The extent to which the landholding classes of late medieval England used writing for utilitarian purposes is exemplified by the most studied family of the later Middle Ages, the Pastons. In addition to literary works, the books owned by the Pastons also contained practical works such as “a ‘little book of physic,’” and “a ‘Great Book,’ containing the *Coronation and the Duties of Knighthood*, *Treatise on War*, a discourse *On Wisdom*, and the *Rules of Chivalry*”, as is revealed by the account of a scribe called William Ebesham, whom Sir John Paston had hired to copy books or him (Thompson 1939: 409). Other non-literary works owned by members of the family included “a number of religious and didactic works; some chronicles and books on heraldry; and a few classics”, as well as a second edition of *The Game and Play of Chess* printed by Caxton (410). Thompson (1939: 409) sees the evidence of their correspondence to indicate that all of the adults in the family “could read and write”, the men, at least, knew some Latin, and some members of the family also knew some French. However, in his edition of a selection of the Paston letters, Davis (1958) provides a somewhat more conservative estimate. According to him, all of the Paston men—although they occasionally employed clerks to write their letters—could most likely both read and write “with differing degrees of competence and elegance” (xxxvi), but the women of the family do not seem to have written any of their letters themselves, instead relying on “whatever literate person happened to be most readily at hand” (xxxvii), as is indicated by the large number of different hands in which their letters are written.

## 6.5 Conclusion

The medieval recipe collection as a written genre emerged at the same time as both the written word and the vernacular English language were asserting themselves as indispensable aspects of everyday life. The spread of literacy down the social scale and the increasing use of English for written communication served to broaden the potential readership of all kinds of texts and also created a competitive situation between the old aristocracy and the rising middle classes, as will be seen in chapter 7. 15<sup>th</sup>-century England was a society with a long history of trilingualism that was still in the process of vernacularization, English being a very recent introduction in several walks of life. Its literate culture was strongly influenced by both Latin and Anglo-Norman French, whose vestiges were still firmly ensconced in fields like science, religion, administration, law and international trade. The diachronic development of the linguistic situation of medieval England is aptly summarized by the following “thumbnail-sketch” by Schendl and Wright on the language used for English civic documents:

[P]rior to 1066, civic documents were written in Old English. By the 1070s they were written in Latin, but a form of Latin that was informed by Anglo-Norman French and Middle English. By 1200 a written form of Anglo-Norman had evolved (earlier, one might note, than on the Continent) and by the late fourteenth century, an orderly mix of the three languages was routinely used for records-keeping nationwide [...]. All this changed in the fifteenth century when this trilingual system was abandoned and written proto-Standard English commenced, with written Anglo-Norman continuing only vestigially thereafter in legal writing. (Schendl and Wright 2011: 19)

This multilingual heritage means that the use of multiple languages within a text, which also characterizes medieval recipes, was in no way exceptional but merely a reflection of contemporary literate culture. 15<sup>th</sup>-century England was also a more literate society than it had been ever before or would be for a long time—at least some literacy practices being employed even quite low down the social scale. The aristocracy and gentry were mostly literate—at least to some degree—in several languages, and the rising middle classes, merchants and craftsmen, were making extensive use of various kinds of written documents and even beginning to read for their edification and enjoyment:

Although there remains a common perception of massive illiteracy (generally exaggerated: it was perhaps higher in the early Modern period), medieval culture was considerably more ‘literary’ than even the modern world of a half century ago. (Hanna 2011: 172)



## Chapter 7

# Historical context of Middle-English recipes

It has become increasingly clear, however, that almost every issue that has made theory a matter of controversy, crisis, and polemical debate hinges on the representational status of texts. While texts are presumably instruments of communication they are also institutional facts and cultural interventions that may affect one's sense of personal, religious, cultural, ethnic or national identity, just as they shape historical cultures in manifold ways. To put a complex point simply, the study of texts cannot be cordoned off from the study of culture.  
(Searle 2004: 3)

As was argued in section 2.2, all texts are produced in and shaped by a definite historical context, which provides a frame of reference for the encoding of knowledge in the text. This also means that in order to successfully decode the text and understand the resultant work, we need to have at least some idea of the frame of reference used to encode it. In the case of the *Potage Dyvers* family of recipe collections, this historical context is that of late medieval England of the 14<sup>th</sup> and 15<sup>th</sup> centuries and that part of its society which could afford to take into account culinary niceties beyond the nutritional value and ready availability of food. This chapter, together with chapter 8, is intended to provide the user of the edition with an idea of the context in which medieval culinary recipes were most likely produced and used, and with “a broad understanding of the subject matter” (Edwards and Moffat 1998: 219) of medieval culinary recipes.<sup>1</sup>

---

<sup>1</sup> These general background essays also serve the practical purpose of allowing the explanatory notes included in the edition itself to focus on more detailed technical information and to reference specific sections in these chapters for more general issues, as recommended by Edwards and Moffat (1998: 219).

## 7.1 Cultural context: thought-styles regarding food

In making comparisons between medieval cuisine and that of our own time, it must be kept in mind that whatever differences we find between them are not entirely due to the differences in available foodstuffs and methods of preparation. Since food is so intimately connected to the human body, “it almost inevitably becomes linked with attitudes toward the body” (Sponsler 2001: 6), including the various socially and cosmologically motivated anxieties about bodily borders that have been examined in the classic anthropological study of Mary Douglas 1979. Much of what seems alien to us in medieval cooking is in fact the result of quite fundamental differences in the ideas about food and its relationship to the human body and soul.

Perhaps the most fundamental of these were the scientific conceptions about the nature of the human body and the influence of nutrition on its condition. While the careful attention paid to a healthy diet that characterized medieval medicine may seem surprisingly enlightened to modern readers, it should be pointed out that the foods actually recommended as healthy “were largely determined by textual tradition and were not always those now thought of as especially conducive to health” (Siraisi 1990: 121). In addition to the human body, the consumption of food was also seen as having spiritual implications for the human soul, which meant that the Christian church also imposed various religiously motivated restrictions on food and eating—mainly in the form of the regular periods of abstinence—which played an important role in shaping the medieval culinary tradition. In addition to these formally codified doctrines meant to protect the body and the soul, there also existed a variety of considerations having to do with the highly hierarchical social system. Since food and eating are thoroughly social phenomena, used to express group identity and to define one’s relationship with other groups and with the gods (Mennell, Murcott and Otterloo 1992: 33), they are to a significant degree influenced by the surrounding social system and its values. In the context of the medieval aristocracy concepts such as wealth, prestige and gentility played a significant part in the ways food was prepared and consumed.

### 7.1.1 Medical and dietary theories

Medicine is the science by which the dispositions of the human body are known so that whatever is necessary is removed or healed by it, in order that health should be preserved, or if absent, recovered.

(*Canon of Avicenna*, translated in Siraisi 1990: 78)

The above definition of medicine comes from the Latin translation of the *Qanun* of the Arab philosopher and physician Ibn Sina—or Avicenna, as he was known in medieval Europe—by Gerard of Cremona, known as the *Canon medicinae*. This work, which was based mainly on the Galenic tradition of medicine and the theories of Hippocrates and Aristotle, was used as one of the standard reference works in medical schools all over Europe from the 13th century onwards



(Weiss Adamson 1995: 40, 65). By the late Middle Ages, the principles of the Hippocratic and Galenic tradition of medicine had become common currency at all levels of the medical profession, and had been “assimilated, albeit with varying degrees of understanding, across a broad spectrum of society” (Rawcliffe 1995: 32), penetrating even ‘popular’ medicine in its various forms (Siraisi 1990: 141, 187). This leads Rawcliffe (1995) to estimate that even the average layman had some conception of the medical theory of humors, and “that the more affluent, book-owning classes were quite well informed about the way their bodies worked (or were believed to work)” (43).

Medieval medical theory was based on the Classical concept of the human body as “a microcosm of the universe, [that] functioned in exactly the same way as the universe itself, sharing the same components and responding with great sensitivity to environmental and planetary influences” (Rawcliffe 1995: 32–33). Since recovering an absent health was an uncertain endeavour at best, and often highly perilous (Rawcliffe 1995: 36–37), medieval medicine focused strongly on *preventive* treatment. Thus the medieval physician was supposed to maintain the health of his patients by tailoring their diet, exercise, rest and environmental conditions to suit their personal properties and situation in life (Siraisi 1990: 120). Medieval physicians distinguished between three basic classes of phenomena, which together determined the health or sickness of an individual. The first were the *naturals*, which included the humors, elements and qualities which made up the human body. The second were the *contra-naturals*, which included all kinds of pathological conditions that were inimical to human life. The third class comprised the *non-naturals*, or the “agents necessary to life”, which were loosely grouped under the categories of environment, motion, nourishment, sleep, evacuation and mental equilibrium. Each of these factors could either harm or strengthen the individual, depending on how they were employed (Rawcliffe 1995: 37–40). The proper employment and control of these environmental factors had a whole genre of literature dedicated to it: the *regimen sanitatis*. In its prototypical form, a *regimen sanitatis* consisted of six chapters or parts, each of which addressed one of the six environmental influences: 1) *aer*, 2) *motus et quies*, 3) *cibus et potus*, 4) *somnus et vigilia*, 5) *repletio et evacuatio*, and 6) *accidentia animi*. Of these, the part on *cibus et potus*, discussing matters of nutrition, seems to have been clearly predominant (Weiss-Amer 1992: 71).<sup>2</sup>

Given the central role of food and drink in the maintenance of a healthy regimen, it is not surprising that the learned knowledge of the physician had an important place at the dining board of any significant magnate.<sup>3</sup> In order to maintain his master in health, the physician would prescribe him a suitable daily and seasonal regimen that covered all aspects of his daily life from the frequency of sexual intercourse to the orientation of his house, but in particular specified the foods he should eat, given his natural temperament (Scully 1992: 185).<sup>4</sup> This—as

<sup>2</sup> Among the 23 *regimina* analysed by Weiss-Amer (1992: 71), it was the only chapter that was present in every regimen and accounted for more than half of the total contents of the *regimina*.

<sup>3</sup> For example the doctor of Edward IV of England was supposed to be present at all of the king's meals to inform him of the diet that conformed most closely to the recognized rules (Hammond 1993: 100). Also the 14th century English recipe collection known as *Forme of Cury* indicates in its preface that “it was compiled by assent and auysement of Maisters and [i.e., of] phisik and of philosophie” attached to the court of Richard II (Hieatt and Butler 1985: 20).

<sup>4</sup> The practical involvement of medical practitioners in ‘nutritional design’ and food preparation was

BLOOD	red and sweet, hot and wet, air, heart, spring, childhood, continuous fever, morning, serene/unruffled, Sanguine, Apostle Mark, Jupiter, Lydian Mode, Gemini/Taurus/Aries
YELLOW BILE	bitter, hot and dry, fire, liver, summer, adolescence, tertian fever, male principle, noon, bold/exuberant, Choleric, Apostle Paul, Mars, Phrygian Mode, Virgo/Leo/Cancer
PHLEGM	white and salty, cold and wet, water, brain, winter, old age, quotidian fever, female principle, evening, idle/foolish, Phlegmatic, Apostle Peter, Moon, Dorian Mode, Pisces/Aquarius/Capricorn
BLACK BILE	sour, cold and dry, earth, spleen, fall, adulthood, quartan fever, afternoon, stubborn/insolent, Melancholic, Apostle John, Saturn, Mixolydian Mode, Libra/Scorpio/Sagittarius

Table 7.1: *Medieval associations of the humors.*

Hammond (1993: 100) points out—does not, however, mean that his advice was always heeded; some of the rules formulated by physicians seem to have been overlooked by their wealthy patrons in their desire for variety and “in order to impress both their guests and their own palates” (Scully 1992: 135).

Humoral theory

Medieval medicine was based on the idea of balance, which in turn was defined by the humoral theory. Its roots go back to the Pre-Socratic philosophers of the 6<sup>th</sup> century BCE, and to the theory of the four elements: fire, water, air and earth. The elemental theory was elaborated by subsequent thinkers, who added to it the idea of the four basic qualities: hot, cold, moist and dry, which were in the time of Hippocrates connected with the four bodily fluids—*blood*, *phlegm*, *bile* (also termed ‘choler’, or ‘red’ or ‘yellow bile’), and *black bile* (‘melancholy’)—which form the basis of *humoral pathology* (Siraisi 1990: 104-5; Weiss Adamson 1995: 10-1). These humors, which were considered to be actual bodily fluids, were assigned “largely hypothetical origins, sites and functions” in the human body (Siraisi 1990: 105).<sup>5</sup> By the Middle Ages they were associated with various properties, entities and phenomena, following the Classical concept of the human body as “a microcosm of the universe” (Rawcliffe 1995: 32-33), summarized by Weiss Adamson (1995: 14-15) as shown in *Table 7.1*.

The concept of an individual temperament or *complexio*<sup>6</sup>—the balance of the principal qualities which was the result of a mixture of the four humors in the individual’s body—was central to the medieval medical and especially dietetic thinking. According to the complexion theory, every person was endowed with an idiosyncratic innate *complexion*, an essential characteristic that was acquired at

well established and continued at least into the 15<sup>th</sup> century, as is demonstrated by a collection of 141 culinary recipes in Latin that were collected from different sources by an Italian physician identifying himself as “N. a doctor of Assisi”. According to him, the recipes were intended “to maintain the body in health and in good appetite and relish in accord with the proper times for all foods as required of all the faithful” (Scully 1992: 43-44).

<sup>5</sup> *Phlegm*, which was mostly associated with the brain, was the general term for more or less any colorless or whitish secretion (except for milk and semen), and could have differing characteristics. *Yellow bile* was identified as the fluid found in the gall bladder and was said to be manufactured in the liver, along with *black bile*, which was considered to be stored in the spleen. The liver was also thought to be the producer of *blood*, which occupied a special place among the humors, serving to feed the different parts of the body (Siraisi 1990: 105).

<sup>6</sup> This was the Latin term used for the Greek term *κράσις* (*krāsis*, ‘mixture’) from the 12<sup>th</sup> century onwards.

the moment of conception and persisted throughout a person's life, characterising the person in relation to others. Ideally, an individual's humoral balance was supposed to remain as close to *temperate* as possible, but the exact meaning of temperance in absolute terms varied from person to person and depended on the individual's innate complexion. In addition to personal differences, a person's complexion varied according to conditions of life, external circumstances, age, and sex (Siraisi 1990: 101-3, 106).<sup>7</sup>

For medieval physicians, complexion theory acted as a system of explanation which provided a rational link between sickness and its cure: sickness occurred when the balance of qualities in the body was disturbed, and its cure could be effected by prescribing medications or treatments whose qualities were inversely matched to the patient's disordered complexion (Siraisi 1990: 102). The therapy prescribed by the doctor customarily consisted of "a combination of dietary measures and medication (often barely distinguishable from each other), accompanied by purgation in the form of laxatives, clysters, diuretics, phlebotomy, cautery, fumigation, hot baths or cupping" (Rawcliffe 1995: 53). While the immediate causes of most illnesses were attributed to shifts in the patient's complexional balance, these shifts themselves could be occasioned by various factors: harmful changes in the non-naturals (especially food, drink, air and water) or celestial influences (Siraisi 1990: 123). Since most serious disorders were not conceived of simply as an excess or deficit of some particular humor, but as a result of "a complex interplay of physical and environmental factors", humoral theory provided a flexible—yet hopelessly complicated—explanatory framework that survived well beyond the Middle Ages (Rawcliffe 1995: 46).

### The human body and its functions

The centrality of food in the medical thinking of the Middle Ages is also reflected in the interest of medieval physicians in the mechanisms of illness and in the ways "the body changed or was changed by food and medicines" (Siraisi 1990: 100). In order to understand the role of food in medieval medical thinking, it is necessary to examine the medieval views regarding the functioning of the human body. These views were a result of the interaction between popular beliefs, religious doctrines, medical ideas and Aristotelian philosophy (Siraisi 1990: 113), the main influence being the Galenic physiological tradition.<sup>8</sup> It not only provided the medical community with a richly detailed account of the human body, but also conformed to their empirical observations in many respects.

The body, its internal organs and the four humoral fluids were considered to be divided into three distinct systems, to each of which were assigned specific virtues, operations and faculties. The *vital virtue* was seated in the heart and was

<sup>7</sup> Young people were hotter and moister than old ones (since complexional heat and moisture in both humans and animals were believed to be used up during the lifespan) and women as a whole were colder and moister than men. Complexion also varied according to geographical region: people living in warmer regions were considered to be warmer than those living in colder regions (Siraisi 1990: 101-3, 106).

<sup>8</sup> "By around 1300, some of Galen's fundamental expositions of physiology were being intensively studied in Latin translation at Montpellier, Paris, Bologna, and probably elsewhere. For example, *On Complexions* provided an account of the theory of temperament, and *On the Natural Faculties* explained nutrition, growth and reproduction" (Siraisi 1990: 84).

responsible for the maintenance of life, conveyed in the body by the *spiritus*.<sup>9</sup> The *animal virtues*—the psychic powers of man—were governed by the brain, and were responsible for mental activity, motion and sensation.<sup>10</sup> The *natural virtues*, governed by the liver, included the powers of nutrition, growth and reproduction. Of these, the powers of nutrition and growth were associated with the digestive organs and the veins that disseminated the digested food in the form of blood (Siraisi 1990: 107–108). In addition to being ordered on this scale of ‘nobility’ or ‘spirituality’ (with the heart at the top of the hierarchy), the parts of the body were also valued on a scale of usefulness. On this scale, the highest-ranking part was often the stomach, which as the first link of the nutritive chain was considered to be critical for survival (Rawcliffe 1995: 44). In medieval medicine the stomach was seen as a kind of ‘flexible cooking pot’ or furnace, and the verb used to express the idea of digestion was that of ‘cooking’ (Scully 1992: 127).<sup>11</sup> Like in any other form of cooking, different foodstuffs were also considered to take varying amounts of time to cook and to react differently to the cooking procedure, some becoming runny and passing easily through the body, others toughening up and stopping the digestive tract (Scully 1992: 127).<sup>12</sup>

It was to a great degree because of these conceptions of the stomach that medieval medical theorists were so interested in the composition of meals, although the same theoretical principles seem to have been used to support quite contrary practices. One view held that light food should be eaten first, since it would be digested quickly, making way for the heavier foods, while consuming slowly digested food at the beginning of a meal would have meant that anything consumed after it would have to wait for it to be cooked before being able to leave the body, thus running the risk of getting overcooked and producing corrupt humors (Scully 1992: 127–128). Another view saw it as natural to take the most substantial foods—requiring both more time and the greater heat of the lower part of the stomach for their cooking—first, and saving the richer and more delicate for later in the meal, so that they can cook more gently in the upper part of the stomach (Hieatt and Butler 1985: 5; Lehmann 2003: 28).<sup>13</sup> These interpretational differences led to a fierce debate among medieval physicians over whether light or solid food should be served first (Strong 2002: 110), each side emphasising a different aspect of di-

<sup>9</sup> The organs associated with this virtue—manifested in the heartbeat, the pulse and the respiration—were the thoracic cavity and the arteries, which distributed *spiritus* throughout the body.

<sup>10</sup> These virtues were so named due to their association with the functions of the soul, *anima*.

<sup>11</sup> According to Avicenna, ingested food was first transformed into *chyle* in the stomach, then transported to the liver and finally concocted (literally ‘cooked’) there into blood, the two biles and phlegm. Various stages of concoction purified the blood of superfluities which were excreted, part of the blood ultimately being refined into semen. Most of the blood, however, was used up in nourishing the various parts of the body (Siraisi 1990: 106).

<sup>12</sup> Among foodstuffs of easy digestion were cabbage, lettuce, portulaca (purslane) and other herbs, all moist fruits (especially peaches) and light meats such as chicks, chicken and goat-kid. Also peas and beans were included in this category, and the flatulence they were known to cause was seen specifically as the result of their quick passing through the stomach. Among prepared dishes, most broths and pottages belonged to this category due to their warm and moist nature. On the other hand, foods that were considered to be difficult to digest included all so-called ‘heavy’ fruits, such as chestnuts and pears, and heavy meats such as beef and pork (Scully 1992: 122–34).

<sup>13</sup> For the same reason, physicians cautioned that foods of diverse substances (eg. chicken, fish, beef and pork) should never be combined, since their differing rates of digestion would cause the more quickly digested matter to draw with it undigested and therefore harmful humors from the slower-cooking ones, or would itself be overcooked (Scully 1992: 135).

gestion. Judging from the menus included in the *Potage Dyvers* manuscripts and those examined by Hieatt and Butler (1985: 4–5), the English custom seems to have leaned towards serving substantial foods first.<sup>14</sup> Since heat was the essential resource of the stomach-as-furnace, Bartolomeo Sacchi (1421–1481)—a 15th-century Italian food writer, historian and a librarian to Pope Sixtus IV, better known as Platina—counsels that after-dinner activity be limited to light humour and games, lest “activity and excitement of the mind” draw the natural heat from the stomach, thus rendering it too weak for digestion” (Milham 1998: 109–11). For the same reason, drinking of water—cold and moist by nature—during or after dinner was strongly discouraged by medical authorities (Manzalaoui 1977: 53), while wine—being hot in nature—was considered to aid the digestion.

Another aspect of eating which was frequently discussed in medieval health handbooks was the need to restrict one’s natural appetite. Maintaining moderation in eating and drinking was important, because overeating—defined by Aldobrandino of Siena simply as continuing to eat past the stage at which one still has the appetite to eat more (Scully 1992: 181)—would disrupt the digestive process and produce either unhealthy quantities of one particular humour or “dangerous fumes likely to generate disease” (Rawcliffe 1995: 39). Therefore the general rule given by most physicians was to eat only when hungry, and to rise from the table while not yet completely full (Scully 1992: 181):

be ware of to moche etyng, be the mete never so good, and withdraw the hand, while he hath appetit and desire to ete. For of superfluyte of mete is the stomak stopped, the body greved, the inwitte hurt, and the mete vndigested abideth in the bothom noyovs, and vndified.  
(Manzalaoui 1977: 53)

In the late Middle Ages it was also believed that the nobility were physically more delicate and refined than the coarse and common workers, and thus required a different kind of diet, easier for the digestive system (Scully 1992: 192):

For grosse metes and stronge ben good to a stronge and an hote stomak, for it fareth as a stronge fire, that hath myght to brenne grete wode. But whan the stomak is cold and feble, than vse he sotell and light metes, for that stomak is likened to a fire that brenneth but reedis, lockers and sotell wode.  
(Manzalaoui 1977: 50–51)

### Therapeutic properties of food

In most cultures in earlier times no sharp boundary existed between foods and drugs. The way to health was a correct diet and lifestyle,

<sup>14</sup> This rationale is explained in the rhymed 15<sup>th</sup>-century cookery collection *Liber Cure Cocorum*, which provides the following rule:

For a comyn rule in cure. / Now tas þys for a rewle fulle gode, / All hole futed fuylle in flud / Gose before, and ay þou take / Þo grettis fyrst, savun gose and drake, / Bothe of towne and of toþer, / Also bakyn mete, my der brother, / And most daynté, come byhynde: / Þys is a rewle mad in kynde.  
(Morris 1862: 55)

adjusted to one’s constitution and to the climate. ‘An apple a day keeps the doctor away’—and keeping the doctor away was much more advisable in those days than it is now [...]. (Dalby 2000: 16)

The sentiment expressed by Dalby is especially characteristic of the European Middle Ages.<sup>15</sup> Both Haly Abbas in his *Liber Pantegni* and Averroës make a three-fold differentiation between *food*, *medicine* and *medicinal food* (Weiss Adamson 1995: 17, 79).<sup>16</sup> Food was further categorized according to various criteria, the most influential being the one created by Haly Abbas in his *Liber Pantegni*. His categorization—used with only minor modifications throughout the Middle Ages—divides foodstuffs into groups based on their origins as shown in Figure 7.1.<sup>17</sup>

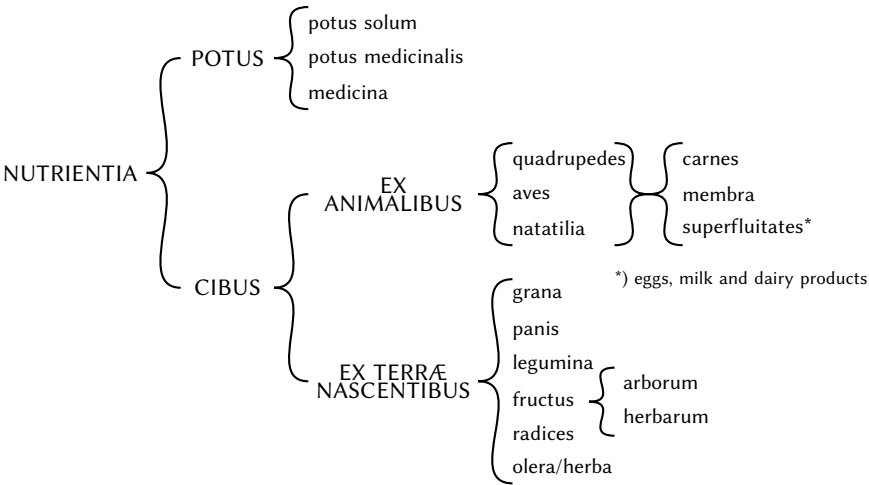


Figure 7.1: Medieval food groups after Haly Abbas (after Weiss Adamson 1995: 196).

In addition to the varying rate of digestion of different foods discussed above, the medieval dietician or chef was also concerned with the *humoral properties* of different foods in terms of the four qualities of hot, cold, moist and dry. Most medieval *regimina*—and also the more compact *tacuina sanitatis*—included this information for the foodstuffs they discussed. Table 7.2, based on a 14<sup>th</sup> century Latin

<sup>15</sup> The overlap in the categories of food and medicine in the Middle Ages, noted by Siraisi (1990: 123), is evident in the explicit of the 14<sup>th</sup>-century English recipe collection known as *Diversa Servicia* which reads: “Explicit de coquina que est optima medicina” (Hieatt and Butler 1985: 79).

<sup>16</sup> However, the categorization of a given item in these categories was not stable. For example confections were originally a part of curative, not preventive medicine, but as Magninus Mediolanensis points out, they had already by the early 14<sup>th</sup> century become viewed as culinary delicacies: “Sciendum igitur quod homines communiter vtuntur in sanitate eorum confectionibus magis ad voluptatem quam propter necessitam” (Scully 1988: 22; Weiss Adamson 1995: 123, 202). Similarly, many of the foodstuffs which throughout the Middle Ages played a central role in preventive medicine (e.g. almonds, sugar, rice, and a wide a variety of herbs, spices and sauces) have since moved firmly into the culinary sphere (Weiss Adamson 1995: 204).

<sup>17</sup> Another categorization, formulated in the *Isagoge* of Johannitius distinguishes between *cibus bonus*, *cibus malus*, *cibus gravis* and *cibus levis*, based on the nutritional properties of the foodstuffs (Weiss Adamson 1995: 41).

version of the *Tacuinum sanitatis in medicina*, translated from the original of the Arab scholar Ellbochasim of Baldach, gives some examples of this categorization.

	WET	DRY
HOT	almond oil, asparagus, aubergine, bananas, beans, butter, cane sugar, celery, dates, fat and lard, figs, grapes, hens' egg (yolk), kid, licorice, liver, mutton, nutmeg, olive oil (black), onions, parsnips, partridge, peacocks, pheasants, pigeons, quail, raisins, roast meats, rocket and watercress, small birds and thrushes, sugar, swedes, sweet pomegranates, tagliatelle (noodles), testicles, turnips, turtledoves, udder, veal, walnuts, wheat	almonds, ambergris, anise, basil, bay tree berries, beetroot, cabbages, carob beans, chestnuts, citrine wine, cow and camel meat, crane, dill, dried and salt meat, dried figs, fennel, galingale, garlic, gazelle, hare, heart, honey, horseradish, hyssop, leeks, lily, marjoram, mint, musk, mustard, old aromatic wine, parsley, pellitory, pickled fish, pine nuts, robust red wine, roosters, rose water, rue, saffron, sage, salt, salt water, salted fish, sweet dates, white wine, wormwood
COLD	apricots, bitter oranges, boiled wheat, brains, cherries, chickpeas, cucumbers, fish, fresh cheese, hens' egg (whites), junket, lamprey, lettuce, melons, peaches, pears, plums, pork, prawns, pumpkins, rainwater, ricotta, sour milk, sour pomegranates, spinach, spring water, truffles, vetchling, violets, watermelon	barley, barley water, black cherries, broad beans, citrons, cream of barley soup, wild dates, gelatin, lemons, matured cheese, medlars, millet, olive oil (green), quinces, rice, roses, rye, sorghum, sour apples, starch, tripe, verjuice, vinegar

Figure 7.2: Humoral properties of some foodstuffs (after Spencer 1984: 140-1).

This division of foods (as well as other substances) into categories according to their properties was further elaborated by differentiating each of the four basic qualities—hot, cold, wet and dry—into four degrees of intensity (*gradus*), enabling a very detailed differentiation between the properties of different substances. This distinction was first attested in written form in the writings of Galen, but he does not yet apply them to foodstuffs. The Latin translation of Haly Abbas’ *Liber Pantegni* describes the degrees from the strongest to weakest as follows: “*magne*”, “*infra hanc*”, “*nec fortes nec debiles*” and “*paucae et debiles*” and expands their range to cover all vegetable foods (Weiss Adamson 1995: 16-7). It was the influential *Tacuinum Sanitatis* of Ibn Butlan that extended the scope of the *gradus* system beyond the vegetable world to cover also meat and animal products, along with the four seasons, different types of clothes and the winds (Weiss Adamson 1995: 91, 200).

Of the categories presented in Figure 7.2, the category of hot and wet foods was considered to be the most universally beneficial, since it corresponded well with the ideal temperament of the healthy human body, considered to be slightly hot and moist. A prime example of an especially beneficial foodstuffs was sugar, which was considered to be identical in temperament to the human body (Scully 1992: 52, 189).<sup>18</sup> Another warm and moist foodstuff considered universally good

<sup>18</sup> Partly due to its increased availability in the 15<sup>th</sup> century and partly due to the virtues attributed to it by Arabic scholars, it became one of the most valued and common ingredients in sick-dishes.

and especially useful for the sick was chicken, which is, curiously enough, omitted from the *Tacuinum sanitatis*.<sup>19</sup> Also wine, considered warm and dry, was seen as beneficial—especially during a meal—since its warming nature served to fortify the digestive capabilities of the stomach.<sup>20</sup> The positive effects of wine were numerous and there seem to have existed no sound arguments against its moderate consumption by healthy adults (Scully 1992: 139).

Also animal meats were considered to be hot to varying degrees, but their moistness or dryness depended on several factors. First of all, different animals naturally had different qualities, pork being the moistest and beef the driest. As a rule, younger animals were considered to be moister than older ones that had dried with age, just as humans were considered to do. Wild animals were also considered to be warmer and drier than their domestic counterparts, since they received more exercise and were more exposed to the sun (Scully 1992: 47). Fish, on the other hand, were by natural association considered to be of cold and wet nature.<sup>21</sup> This meant that both their method of preparation and the condiments served with them were required to heat them up (Scully 1992: 75).<sup>22</sup> Eels and lampreys were held to be especially cold and moist and required even stronger treatment than other fish. According to the *Menagier de Paris*, a late-14<sup>th</sup>-century French guidebook on running a household, they should be killed in a bath of salt (considered very hot and dry) and left there for three days in order to dry and warm them.<sup>23</sup>

As can be seen in *Figure 7.2*, also most fruits were considered very moist, and were therefore supposed to be eaten only when roasted or baked, or combined with dry ingredients (Scully 1992: 45).<sup>24</sup> Most cereal grains were of a moderately cool and dry nature, and were therefore especially suitable for porridges,<sup>25</sup>

---

It was also an important component in many medical preparations, such as syrups, theriacs and electuaries (Scully 1992: 52, 189).

<sup>19</sup> According to Magninus Mediolanensis, young hens, young roosters and their pullets, and fat capons were the most temperate of all birds, and thus a staple ingredient of sick-dishes (Scully 1992: 188).

<sup>20</sup> Its popularity as a cooking liquid and as a base for medicines was further enhanced by it being considered to be very ‘subtle’ (i.e. it vaporized easily) and to transfer the humoral properties of food particles better than water (Scully 1992: 138).

<sup>21</sup> The best (i.e. least harmful) of them were those which most resembled land animals: porpoise, shark, dolphin and cod. In the *regimen* of Magninus Mediolanensis these were followed by the intermediate category containing fish such as red mullet, gurnard, plaice, sole and whiting. Near the bottom of the list appear salmon, turbot, mackerel and conger, which were not recommended due to their excessive moistness (Scully 1992: 49).

<sup>22</sup> For example recipes *PD 132* and *PD 168* in the *Potage Dyvers* collection indicate roasting as the first step in preparing salmon, and instruct to serve it in heating and drying sauces (vinegar and parsley in *PD 132*; wine and vinegar sauce with ginger and cinnamon in *PD 168*).

<sup>23</sup> Although eels were never supposed to be boiled, but either fried or roasted and served with a hot and dry sauce—often containing red wine, salt and black pepper, which were considered to be extremely hot and dry (Scully 1992: 76)—recipes *PD 25* and *PD 34* instruct to boil the eel, apparently counting on the power of the hot and dry spices to counter its cold and wet nature. For lamprey, recipes *PD 30*, *PD 60* and *PD 130* provide instructions for roasting, baking and boiling them, although it should be noted that the lamprey to be boiled is described as salted, which would have served to dry and heat it up.

<sup>24</sup> For example raw pears were considered by medical authorities to be so cold and moist as to be poisonous; even in cooked state, they should always be eaten with wine to dry and warm them (Scully 1992: 45). For examples of this, see recipes *PD 5*, *PD 53*, *PD 57*, and *PD 166* in the *Potage Dyvers* collection.

<sup>25</sup> For example barley broth or barley water was considered to be an ideal sick-dish, being slightly warm and moist as a result of the combination of the ingredient and the cooking method (Scully



although wheat was considered temperately warm and dry and therefore used for a particularly wide variety of functions (Scully 1992: 68).

As a logical consequence of these properties, there also emerged a tradition of habitually serving certain dishes or foodstuffs together for medical reasons. For example any fish served in a meal was always supposed to be followed by nuts, which were by nature dry and would therefore absorb the superfluous humidity of the fish. Similarly, meat was habitually followed by cheese, which had the property of quickly descending to the bottom of the stomach, where digestion was most effective, 'pressing' all previously consumed food under it and facilitating its digestion (Scully 1992: 134).<sup>26</sup>

In making up his daily menus, the medieval cook, however, had to—at least in theory—match the humoral properties of the foods not only to each other and their cooking methods, but also to the individual temperaments of his patrons and to the season of the year. The medically motivated distinction between dishes suitable for winter and for summer was first recognized by Magninus Mediolanensis in his regimen, which also presented a method for adapting dishes to the season: whereas dishes eaten in the winter should have wine added to them, vinegar or verjuice should be used during the summer (Weiss-Amer 1992: 76; Weiss Adamson 1995: 202). The personal differences between people (see above) meant that what was good for one person could be quite harmful to someone else (Scully 1992: 101–102). For example in the regimen of Petrus Fagarola, written for his sons, he cautions them of raw onions and undiluted wine—of which the latter was generally considered very healthy, especially in winter and for older people—due to their extreme heat (wine) and moisture (onions), which could be dangerous to young people who tend to excessive heat and moistness (Weiss Adamson 1995: 120).

Looking at the variety of medical rules and recommendations concerning food, it is hard to avoid being struck by the simultaneous alienness of the theoretical reasoning behind them and the eminent practicality of many of its conclusions. Although the concept of the humors or the idea of blood as the fuel of the body have since fallen by the wayside, the efficiency of sugar as a source of nutrition (much to the detriment of modern man) or the beneficial effects of wine enjoyed in moderation are still accepted today as they were six hundred years ago. This contrast between the practicality of the results and the often fanciful nature of the theories are apt to raise suspicions as to which came first. For example Scully (1992: 51) considers it justifiable to question whether the rationalizations about the nature of foodstuffs might have been the result of some particularly inspired elaboration of some dish, instead of being the reason for the said elaboration. Also the sheer convenience of many beliefs points to similar inverted causality. For example, the month of September—conveniently coinciding with the period of harvest and plenty—was in medieval health handbooks often considered to be of a nature that is harmful to no-one and permits the consumption of all foods without harm (Scully 1992: 102).

---

1992: 68).

<sup>26</sup> These principles are contained in one of the best-known verses of the *Regimen sanitatis Salernitanum*, located in chapter 40: "Post pisces nux sit: post carnes caseus adsit" (Weiss Adamson 1995: 120). Another traditional rule, known as the '*Regula Ypocratis*', is contained in both Avicenna's *Canon* and the *Secretum secretorum* and states that fish and milk should never be consumed together (Weiss Adamson 1995: 71).

As Scully (1988) points out in the context of the French recipe collection known as the *Viandier de Taillevent* mediaeval recipe collections do not always fully conform to the recommendations made in the dietetic literature of the time as to how they should be cooked or seasoned. However, he argues that for example the *Viandier* does reflect “a culinary practice that recognizes in a general way the doctrines propagated by contemporary medical schools concerning the most wholesome means of cooking and preparing particular meats and the most salubrious condiments to be consumed in conjunction with them” (22-3). Regardless of the origin of the medical and dietetic theories, their relationship to the actual culinary practices as described by surviving recipes remains an interesting question. Like the *Viandier*, the *Potage Dyvers* collection contains examples of both conformance and non-conformance to the medical theories, but the facts that even the theories formulated by medieval dietiticians are not entirely consistent internally and that medieval cooking was influenced by other factors apart from medical considerations (as will be shown below) mean that the relationship of the *Potage Dyvers* and other English recipe collections to these theories is bound to be extremely complicated, and will provide a suitable topic for further study.<sup>27</sup>

### 7.1.2 Religious doctrine

The Christian church played an important role in the shaping of medieval European culinary habits. It inherited from the Judaic tradition the practice of regulating people’s eating, and by the Middle Ages food had already become subject to God-given rules, and thus also a moral issue (Strong 2002: 50). Not all of the influences of Christianity on the medieval table were negative, however. For example the numerous Biblical examples of eating emphasized the role of communal dining as an expression of love, communion and fellowship and sanctioned the barbarian tradition of celebrating all major events with a feast (55): “Even if the food was simple and the home poor, every time the table was laid for meal it was a reflection of the Last Supper” (Paston-Williams 1993: 63). By imposing the same customs and habits on all of western Europe, the Christian church also helped to unify the ways in which people of different nationalities approached food and to give medieval European cuisine a truly international flavour. There were two Christian customs that had an especially strong impact on medieval culinary habits:

Food symbolised many things for medieval Christians. But the most important Christian food practices were fasting and eucharist. Christians male and female paid tribute to God’s power and acknowledged their own sinfulness by renouncing food. And Christians male and female received their God most intimately in that holy meal in which he became bread and wine.

(Bynum 1987: 31)

<sup>27</sup> The inclusion of regularized word forms and part-of-speech information in the present edition (see subsections 10.3.1 and 11.9.1) makes it easy to identify the names of ingredients and cooking procedures in the recipes and to use them, together with a database of humoral properties of foodstuffs and cooking methods, to study the humoral balance of the dishes described by the recipes.

### The significance of fasting

In the spirituality of late medieval Europe, consuming food and abstaining from it were powerful symbols. Fasting was a common form of piety, and there was a wide variety of medieval men and women who “adored Christ in the bread and wine on the altar, received eucharistic visions and worked to propagate eucharistic piety” (Bynum 1987: 73-4).<sup>28</sup> Of these religious food practices, it was the institution of the fast—and the rules and regulations associated with it—that most significantly shaped medieval culinary practice. It must, however, also be kept in mind that the religious significance of food and especially fasting changed significantly from the days of the early church to the late Middle Ages. By the 14<sup>th</sup> century, the communal and consolidating aspect of both fasting and communion had been very much backgrounded, as they became more and more individualistic “matters of attenuation and dispensation” (32).<sup>29</sup> Especially fasting came to be seen less as a display of self-control offered to God as an atonement for the original sin, and more as a way to partake of Christ’s agony and to express the “unquenchable thirst for mystical union” (33). Abstinence and fasting were also coupled with charity and almsgiving from very early on, as Pope Leo the Great wrote in the 5<sup>th</sup> century: “Let the abstinence of the faithful become the nourishment of the poor and let the indigent receive that which others give up” (quoted in Bynum 1987: 31).<sup>30</sup> Many of the patristic writers also made an association between food and lust, and urged abstinence from eating as a method of curbing sexual desire. An example comes from Jerome, who quoted Terence at a widow asking for advice on how to control her desires: “Sine Cerere et Libero friget Venus” (37). Control over eating was thus equated with control over sexuality, and religious writers warned especially women that food was dangerous because it excited lust in the flesh. This association lasted throughout the Middle Ages, and stories of women who fasted in order to quell sexual desire survive from many late medieval hagiographers (214).

It is also important to note that even the influence exerted by the Christian church and its fasting regulations was to some degree based on the very same thinking that lay behind the medical theories of the time (Scully 1992: 58). Christian writers drew upon pagan philosophy as well as the scriptures for support for fasting, and it was the “Pythagorean and neo-Platonic desire to escape the body” that loomed behind the Christian idea of fasting as a way of “moderating lust, cleansing the brain and body, and preparing the soul for God’s inspira-

<sup>28</sup> Although many modern scholars have discussed these phenomena under various medical and psychological rubrics—from rejection of the mother to fear of mutilation—it must be kept in mind that the different kinds of food miracles, forms of eucharistic piety and fasting are first and foremost food practices, and as such intimately connected with the culinary practices of the Middle Ages (Bynum 1987: 75).

<sup>29</sup> In the Antiquity and up to the High Middle Ages, fasting and Sunday eucharist were a strong unifying factor within the church: they were what every Christian had in common. To partake of the communion was to be united with one’s fellow Christians, while to fast—either in preparation for the weekly Sunday meal, in Lenten anticipation of Easter or in seasonal response to the harvest—“was to join with scarcity in order that plenty might come” (Bynum 1987: 33).

<sup>30</sup> Also Saint Augustine connected fasting explicitly to charity as he wrote: “Above all be mindful of the poor so that you lay up in the heavenly treasury what you withhold from yourselves by a more frugal mode of life. The hungry Christ will receive that from which the fasting Christian abstains.” (Quoted in Bynum 1987: 35.)

tion" (Bynum 1987: 36).<sup>31</sup> Despite all the physiological evils that were attributed to immoderation in eating, it was looked upon essentially as a moral question: overeating and failing to practice abstinence on the allotted periods exemplified the cardinal sin of gluttony. This aspiration for moderation was a central concern in all aspects of medieval morality, and also a significant commonality between the medical and religious aspects of food (Scully 1992: 181–182).

### Fast days and feast days

In the Middle Ages, the church with its daily routine of services and its annual calendar of feasts was the principal means of indexing the passage of time. In terms of culinary practice, the most significant distinction was that made between weekly and yearly fast and feast days. This distinction also had considerable theological importance:

Throughout the Middle Ages, the Lenten fasts and weekly fast days, especially Fridays, remained basic marks of the Christian. In the thirteenth and fourteenth centuries a Christian was, as a minimal definition, someone who received yearly communion, fasted on Fridays and in Lent, paid tithes, and had his or her children baptised.

(Bynum 1987: 40)

The earliest fast days, Wednesday and Friday—known as the *stations* and modelled after the Jewish fasts of Monday and Thursday—appeared already during the first centuries of the church, and were later supplemented by Saturday (as an extra day of fast or *superpositio*) (Bynum 1987: 37). The most significant of these in the late medieval week was Friday, which was apparently supposed to be an actual fast day, allowing only a single light meal (although most laymen seem not to have observed this requirement) (Hammond 1993: 19). On the other days of abstinence, Saturday and Wednesday—and sometimes also Monday—only meat was forbidden (Scully 1992: 61).<sup>32</sup>

Of the annual fasts, the first and most important was the *Lent*, or *quadragesima*, covering the 40 days preceding Easter. Later additions included the *Lent of Pentecost* (ending on Peter and Paul's day, June 29), the *Advent* (beginning November 14)<sup>33</sup>, and the so-called 'quarter days' (*quattuor tempora*), which were later called the Ember days in English (Bynum 1987: 37).<sup>34</sup> Although the number of criteria that qualified for exemption from fasting increased in the period from the 13<sup>th</sup> to

<sup>31</sup> For example St. Isidore of Seville, the learned doctor of the church, based his praise of the moral usefulness of fasting on the familiar scientific principles: meat and other foods with a warming effect were hazardous to one's moral health because their warmth awakened the lust of the flesh (Scully 1992: 59).

<sup>32</sup> For example a copy of the *Viandier of Taillevent* from the beginning of the 15<sup>th</sup> century, named Sunday, Tuesday and Thursday as the days when meat was allowed. Thus every Monday, Wednesday, Friday and Saturday was marked by abstinence from red meat (Scully 1992: 61).

<sup>33</sup> This period of abstinence originally developed as an expression of penitence offered at the end of the year, and only later on developed into the period of anticipation we know as Advent (Bynum 1987: 37).

<sup>34</sup> In the liturgical calendar of the church, these days were fixed by Pope Gregory VII (1073–1085) as the Wednesday, Friday and Saturday—the normal fasting days—after December 13 (S. Lucia), after Ash Wednesday, after Whitsunday, and after September 14 (Exaltation of the Cross) (Herbermann 1909).

the 15<sup>th</sup> centuries and preachers started to treat fasting as a symbol rather than a concrete act, some form of abstinence was certainly practised each week throughout the year in late medieval Europe (239). Although the exact days of fasting seem to have varied by region and by century, there had by the late Middle Ages developed an idea that all great feasts and significant high moments in the ecclesiastical calendar should be preceded by fasting (Scully 1992: 60; Bynum 1987: 37).

On the prescribed days of abstinence, meat and by-products of four-footed animals were prohibited. This led to the flesh of birds and fish, which were not explicitly forbidden, becoming the standard fare of these so called 'lean days' (Scully 1992: 74). The inclusion of fowl in the prohibition of red meats on days of abstinence seems to have varied from time to time and place to place (76), but whether they were strictly prohibited or not, it was fish that was most strongly associated with abstinence.<sup>35</sup> As an example of Lenten consumption of fish in a gentry household, we can look at the surviving account books of Dame Alice Bryene of Acton Hall, Suffolk, from 1412–13, which show the consumption of herring rising by about 50 percent, and also fresh fish and shellfish being purchased in greater quantities than usual. This is no wonder, since apparently the household continued to eat three meals a day instead of the one prescribed by the church (Hammond 1993: 74).<sup>36</sup>

The strictures and prescriptions of fasting had wide-ranging consequences for the culinary habits of medieval people. They not only dictated what could be prepared and eaten, but also shaped the 'food industry' of the period as a whole: the annual periods of fasting had a strong impact on the business activities of slaughterhouses and fishermen alike (Scully 1992: 62). The profession whose work the religious restrictions on food impacted most directly, however, was the cook: "In meeting the requirements of lean cookery, the medieval cook became very adept at making substitutions" (Scully 1992: 90).<sup>37</sup> Furthermore, since the task of the professional cook was not limited to producing individual sauces, pottages, and pies, but whole meals, any considerations for a lean day had to begin on the level of the whole menu. It was not merely a question of avoiding or substituting certain foodstuffs, but of designing the whole meal within the given parameters in a manner that would not make it seem inferior to its meat-day counterpart (Scully 1992: 90).<sup>38</sup>

<sup>35</sup> Based on the bills of fare included in the different manuscript versions of the *Potage Dyvers*, it would seem that fowl were not considered appropriate for days of abstinence in 15<sup>th</sup>-century England, as they do not occur in the menus which do not contain red meat. It should, however be pointed out that the definition of fish was a rather loose one: the category also included whale, dolphin, porpoise, beaver's tail and barnacle goose (Wheaton 1983: 12; Bynum 1987: 41).

<sup>36</sup> This seems to have been quite common; even the monks of Winchester Cathedral Priory ate three meals a day during Lent and alleviated the austerity of fasting by eating a large quantity of figs and raisins (Hammond 1993: 74).

<sup>37</sup> An example of this ability can be found in the Viandier of Taillevent, which presents a recipe for a Lenten tart that uses fish roe and milt together with almond milk to approximate the taste and feel of a cheese custard tart. Other substitutes included 'eggs' made from fish roe or almond milk and 'bacon' slices made from fish (Wheaton 1983: 12). Examples of such substitutions can also be found in the *Potage Dyvers* collection (see recipes PD 156 and PD 169).

<sup>38</sup> The bills of fare for various feasts included in the manuscript versions of the *Potage Dyvers* also include menus for lean days, although they are not explicitly labelled as such. For example the fourth menu in MS Ad (f. 24v) and the third and sixth menus in MS H279 (f.47v and 48v) consist

### The ideal versus the reality

As can be inferred from the preceding, the ideal of depriving oneself of the pleasure of food was not always pursued in earnest. According to Bynum (1987: 41), the evidence from medieval cookbooks seems to indicate that in general, the aristocracy observed fasting strictly but legalistically. For example the 15<sup>th</sup>-century household of Dame Alice Bryene seems to have adhered strictly to the pattern of meat on Sunday, Monday, Tuesday and Thursday, and fish on other days (Hammond 1993: 74). Already Thomas Aquinas was aware of this rather legalistic approach to fasting that was gaining hold in the 12<sup>th</sup> and 13<sup>th</sup> centuries and feared that gluttony might be tainting the abstinence of those Christians—some of them even monks—who followed the letter of the fast by abstaining from meat, yet ate lavish feasts prepared from the allowed ingredients. He reminded his readers that eating rich, expensive dishes fell under the compass of greed just as well as eating too much food or eating food prohibited by the church (Bynum 1987: 41). In general, the tone of late medieval preachers and theologians regarding food and fasting seems to have been one of moderate, rational decency: excess was discouraged in either direction (240).<sup>39</sup>

As both Mennell (1986: 28) and Scully (1992: 124–125) point out, it is probable that medieval people accepted the strictures of the fast so easily precisely because their observance in fact demanded very few concessions. The poorer classes probably had little meat available to them outside the feast days in any case, and for the upper classes, the alternatives to a normal meat-day meal were “not really all that bad” (Scully 1992: 124):

[...] there was a certain irony in the way that the Church’s insistence upon “abstinence” led medieval cooks to develop dishes whose gastronomic appeal lost nothing in comparison with those of their creations that were licit on meat days. The cooks’ inventiveness and mastery of their craft must have made “abstinence” quite a bit more enjoyable for the wealthy Christian. (Scully 1992: 91)

Overall, looking at medieval cookbooks one is left with the impression that the medieval aristocracy did in fact manage to dine rather sumptuously even within the strictures of abstinence.<sup>40</sup>

### 7.1.3 Social factors

As Halliday (1978: 60) has pointed out, the surrounding social system is an essential factor in decoding any text. In the case of historical texts, the social context—due to our increased distance from it—is at the same time more salient and less accessible than in the case of contemporary ones. Since “[t]astes in food, like tastes in music, literature or the visual arts, are socially shaped, and the major forces which have shaped them are religions, classes and nations” (Mennell 1986:

---

entirely of fish and other non-meat dishes.

<sup>39</sup> Even St Francis of Assisi—who himself is reported to have practised severe food asceticism, rarely eating cooked food and whenever doing so, adulterating its taste with water or ashes—required the Franciscan brothers to fast only on Fridays, in Advent and in Lent (Bynum 1987: 95).

<sup>40</sup> See for example recipes PD 20, 30–31, 46, 60, 65, 73, 75, 80, 91, 101, 104, 130–165 167–170, 173, 179, 182, and 184 for dishes suitable for days of abstinence.

17), accounting for the social context is especially important in the case of culinary texts. Since religious influences were already discussed above, this subsection will focus on the influence of the sociopolitical sphere, which in the medieval context was centred primarily around the concept of *class*:

In medieval times, differences between the strata of society in matters of food, as in many other aspects of manners, were more striking than differences between countries. At the highest level, a grand banquetting cuisine was common to the great courts of Western Europe.  
(Mennell 1986: 40)

This influence of class is made all the more significant by the dynamic nature of late medieval society: in terms of social structure, the most significant feature of the late Middle Ages was the ongoing struggle between the church, the nobility, and the rising urban bourgeoisie. These struggles and their eventual outcome were also reflected in the “cultural products of society, including how and what people ate” (Mennell 1986: 54).<sup>41</sup>

### Social hierarchy

Perhaps the most important social feature of the culinary culture of late medieval Europe is its association with the concept of *hierarchical man*. Just like everything else in the medieval world from the angels in heaven to the humblest animals, plants and minerals, also mankind was divided into different categories that had their own rights and obligations. Although the division of society into the three estates or *ordo*—*oratores, bellatores, laboratores*—was by the beginning of the 15<sup>th</sup> century increasingly losing its relevance, it was merely giving way to a more complex hierarchy based mainly on wealth and landholding, but also on professional or commercial success (Keen 1990: 2–11). Although the emerging middle classes had not yet found their place in the traditional scheme of things, they did little to bridge the gap between the high and the low. On the contrary, they provoked in the aristocracy a furious attempt to maintain and strengthen the distinction between themselves and the common throng, which was also reflected in food practices (Mennell 1986: 17). According to Freeman (1977: 144–145), it is precisely this type of strictly hierarchical society that is one of the prerequisites in the development of a socially differentiated *high cuisine*. The mentality underlying this development is embodied in the idea, that while the medieval poor ate to live, the wealthy lived to eat (Mead 1931: 9; Goody 1982: 135). In this light, the culinary culture that developed in medieval England and the other great European nations could be described as a *sumptuary cuisine* or a form of hierarchical cooking (Goody 1982: 134):

Since its preparation and consumption had such important implications for hierarchy, food tended to be the subject of competition between those of similar status as well as the subject of regulation between those of different rank.  
(Goody 1982: 140)

<sup>41</sup> For this very reason, also social scientists, among whom Pierre Bourdieu, Claude Lévi-Strauss and Mary Douglas are perhaps the most prominent, have over the past half a century taken an increasing interest in the food habits of different cultures, including medieval Europe (Mennell 1986: 6).

Also Scully (1992: 202–203) believes that there undoubtedly was a great deal of emulation and competition between courts across Europe, as the nobility strove to show first of all their exceptional worth and secondly their homogeneity. No longer a question of life and death, food became an important means of affirming one's status: "In the Middle Ages, as today, there was a lot of snobbery attached to food" (Scully 1992: 202). With the association of a certain type of culinary culture with the upper classes also came the desire of the upwardly mobile middle classes to emulate it. For example the recipes of the French *Ménagier de Paris*, a late 14<sup>th</sup>-century household manual for the urban bourgeois, "resemble those of courtly cookbooks, suggesting an emulation of courtly eating by those in lower social strata", and the same pattern is also seen in "some of the surviving English cooking books from the fifteenth century" (Sponsler 2001: 15).<sup>42</sup> This emulation, in turn, provoked the upper classes to defend the privileged nature of their culinary culture through legislative regulation.

This regulation took the form of what is known as *sumptuary legislation*—laws that were meant to uphold the distinction between classes through various kinds of restrictions. Since these laws were not targeted only at food but also at the clothes and jewellery worn by people, it is clear that their concern was not food consumption as such, but rather food consumption as *social display* (Mennell 1986: 29–30).<sup>43</sup> Legislation was not the only response prompted by the rising threat of the numerous newly ennobled families and the rising wealthy bourgeoisie. According to Strong (2002: 102), the threatened aristocracy also responded by escalating the pomp associated with formal occasions "as a means of preserving caste".<sup>44</sup>

### ***Courtoisie*: the importance of good manners**

As Keen (1990: 160) points out, the worship or standing of a medieval nobleman or a gentleman in the eyes of his peers and neighbours was dependent not only on his lineage, but on the manner in which he conducted himself. The household and its proper operation was an important measure of its master's status, as Georges Chastellain—the chronicler of the 15<sup>th</sup>-century Burgundian court—writes:

After the deeds and exploits of war, which are claims to glory, the household is the first thing which strikes the eye, and that which it is,

<sup>42</sup> Unfortunately, Sponsler does not mention specific examples. None of the *Potage Dyvers* versions show any explicit indication of their intended target audience or its original owner, which makes it difficult to determine whether they were compiled for an upper- or middle-class audience.

<sup>43</sup> In the domain of food these laws attempted to restrict excessive display of wealth in the dinner table, like the proclamation given by Edward II in 1316: "Pur ceo qe par trop outraiouses et desmesurables services de mes et viaundes qe les grantz seignures de nostre reame einz ces heures unt fet et uses de fere, et uncore fount et usent en lur hostel, e de ce qe autres meindre gentz de mesme le realme, a queus teles choses ne appent pas de enprendre, se aforcent de coudre faire les graunts en fesauns teles utrages, outre ce qe lur estat demaunde;" (Stubbs 1882: 238). These laws do not seem to have been very efficient, however, judging from the fact that new, increasingly detailed laws were passed by successive monarchs (Paston-Williams 1993: 67).

<sup>44</sup> The leader in this trend seems to have been the Burgundian court, whose political and cultural authority was in the 15<sup>th</sup> century concentrated in the southern Netherlands with which London had close trade connections (Keene 2000: 99). Therefore it is not surprising that the Burgundian style of dining, along with its style of book production, painting and dramatic spectacle, were subsequently imported to England under Edward IV and Henry VII (Meale 1989: 206; Strong 2002: 102).



therefore, most necessary to conduct and arrange well.

(Translated in Keen 1990: 160)

The measures of a medieval lord's esteem included the number of his dependants, the 'cut of their cloth', and the standard of provision and amplitude of hospitality that his household could extend to visitors (Keen 1990: 160; Paston-Williams 1993: 63). Hospitality was expected of every aristocrat from a humble esquire or gentleman to the mightiest earl or duke, and as a measure of the host's means and worship, it was expected to be lavish.<sup>45</sup> One important vehicle for displaying one's wealth and generosity to one's subjects and to foreign visitors was the institution of the medieval feast (discussed further in subsection 7.2.3 below):<sup>46</sup>

The number of retainers demonstrated to visitors a lord's importance; his status was further enhanced by the degree of ceremony and ritual attending the serving of meals. Great feasts were, above all, occasions for display: the brilliance of the surroundings, the costume of the diners, the liveries of the many servers and the tableware, were matched by an abundance of food based on substantial pottages and good roasts, served on a lavish scale. Feasts held in smaller manor houses were more modest, but ritual and ceremony remained extremely important.

(Paston-Williams 1993: 63)

Although both the food eaten and the manner of its serving expressed the differences in rank among the diners, the grand banquets and the commensality they entailed between their participants also "stressed certain bonds between those sitting above and beyond the salt", serving as a public demonstration and strengthening of the political hierarchy (Goody 1982: 142).<sup>47</sup>

Also the manners of eating—which fundamentally amount to the ability to govern one's basic appetites—were seen as a demonstration of one's relative level of *refinement*, by which one was measured as a social person (Scully 1992: 175). The medieval concept of etiquette was based on the hierarchic principle and on knowing one's place in the social hierarchy: a well-mannered person, regardless of his or her rank, knew the limitations of that rank and acted accordingly (Wheaton 1983: 6). Not much was expected of the lower classes, but anyone who expected to be invited to dine in an aristocratic table, let alone host one, was supposed to conduct himself or herself in a suitably refined manner. Aristocratic youths, whether brought up at their own household or that of a greater magnate, would have learned most of these manners through practical instruction and practice. For example in connection with meals, they would have been expected to "perform certain symbolic acts of service to their fathers or lords: to carve meat, serve

<sup>45</sup> This is exemplified by the description of the "worthy vavasour", Chaucer's franklin: "An housholdere, and that a greet, was he; / Seint Julian he was in his contree. / His breed, his ale, was always after oon; / A bettre envyned man was nowhere noon. / Withoute bake mete was nevere his hous, / Of fissh and flessch, and that so plenteuous / It snewed in his hous of mete and drynke" (Benson 1988: 29).

<sup>46</sup> Wheaton (1983: 3) goes as far to suggest that the demonstration of wealth and power in order to impress was the principal reason for arranging feasts.

<sup>47</sup> This simultaneous emphasis on hierarchy and on the bonds crossing it is reflected in the expectation that the lord (or lady) of the house display his or her generosity by distributing "tidbits of special dishes" to those below him or her at the table, who would not otherwise receive them (Hieatt and Butler 1985: 5).

drinks, hold lights, and bring water for washing the hands” (Orme 1989: 172). In the 13<sup>th</sup> century there also developed a new vernacular genre of instruction books, based on an earlier Latin tradition, which were devoted to establishing the rules of *courtoisie*—of courteous and civilized behaviour (Goody 1982: 143; Orme 1989: 15). Many of these texts place considerable “focus on various aspects of eating, including such things as proper seating order, correct techniques for serving, and decorous ways of ingesting food” (Sponsler 2001: 2), reflecting the importance of the social and symbolic function of food and eating. These works were—at least ostensibly—intended partly for the education of servants and partly for the upbringing of young squires (Hammond 1993: 104), but in the 15<sup>th</sup> century they were often owned—and supposedly used—by the members of the mercantile class, who would have found them useful for learning the virtues of the upper classes that they were aspiring to (Nicholls 1985: 70-1).

During the 14<sup>th</sup> and 15<sup>th</sup> centuries, the copying of these treatises exploded (Nicholls 1985: 157), “demonstrating the spread of literacy among the laity as much as a keen desire to climb the social ladder” (Strong 2002: 112), and indicating an increase in the number of people who “needed—and wanted—to learn the ‘right’ way of doing things” (91). Amos (2001) sees these courtesy books as textual presentations of “behaviors of privilege”, which made “noble education available for appropriation by the powerful common lay readership” (30). Since the “appeal of an aristocratic education for those seeking to enhance their class status” is obvious, “it is unsurprising that historical and literary evidence argues that these encodings of aristocratic *noriture* were eagerly sought after and appropriated by the urban elite” (30):

Fifteenth century society was competitive and fluid, yet wished to maintain respect for order and degree. It is no accident that this century, when the number of “bourgeois gentilhommes” was quickly increasing, saw the appearance of more books on etiquette than ever before, telling one how to behave in courtly circles and devoting special attention to subtle differences and equivalences of rank.

(Myers 1959, quoted in Amos 2001: 25)

Considering the virtues promoted by these etiquette books—including modesty, temperance and respect (Wheaton 1983: 6)—side by side with the stereotypical image of the common medieval peasant, it seems obvious that an important function of these treatises was “to inculcate a code of manners that is specifically contrasted with that of the rustic” (Goody 1982: 140) and to maintain a clear class distinction of manners between the upper and lower classes.<sup>48</sup>

Based on the above, it seems evident that “[p]repared food has long played an important role in proclaiming the social status of an individual or family” (Scully 1992: 118) and “that the great differences which existed between social estates in the centuries of feudalism were reflected in what people ate” (Mennell 1986: 41). These sociological observations also help explain the significance and functions of medieval culinary writing (discussed in chapter 8). Culinary manuscripts,

<sup>48</sup> This is also suggested by the observation that the manual of instruction for good manners seems to be a genre typical to a highly hierarchical society: the European examples are paralleled by Indian, Arabian and Chinese examples from an earlier period, in which all of these cultures shared the hierarchical nature of medieval Europe (Goody 1982: 144).

disseminating recipes from the great noble courts of Europe, provided the “basic standard of excellence” (Scully 1992: 118) required by this distinguishing function of cookery—a model of noble cookery, to which the culinary achievements of an individual household could be compared, and to which the rising classes could both aspire to and be judged by. Simultaneously, they let the aristocrats “know what membership in their classes *entitled* them to by way of prepared food” (Scully 1992: 118). Accepting the opinion of many social scientists, voiced by Mennell (1986: 53), “that people come positively to like foods which developing social standards define as desirable”, we may well surmise that the popularity of the more extravagant dishes among medieval aristocracy was not necessarily due to their culinary superiority, but rather due to their value as social symbols of opulence.<sup>49</sup>

## 7.2 Situational context: the medieval household

After having established the institutional and ideological context of food in the Middle Ages, it is time to step down the ladder of abstraction and briefly outline the more concrete environment in which medieval European cuisine was practised: the households of the late medieval gentry and nobility. The importance of food as a token of social status, discussed above, is highlighted by the fact that a substantial part of the expenditure of an aristocratic or gentry household was devoted to culinary and gastronomic ends, provisioning being almost always the largest item in manorial accounts (Keen 1990: 168-9; Hammond 1993: 63; Scully 1992: 245). It seems that lesser members of the gentry could very well spend more than half of their income on food and drink, while rich knights would spend rather less than half and wealthy earls less than one quarter. Lesser households also spent most of their expenditure on basic staples such as bread and ale, while the greater ones spent relatively little on staples and a greater proportion on luxury articles such as wine and spices (Dyer 1989: 55-6; Hammond 1993: 75-6).

### 7.2.1 The medieval kitchen and its staff

A medieval household of any stature employed a large number of people in a wide variety of tasks, requiring varying degrees of formal and informal training. With regard to the preparation of food, the most important individual member in the workforce was the head cook, followed closely by the physician, whose task was to ensure that his master’s diet remained conducive to his health. But in addition to these two professionals with considerable training of very different kinds, there existed a more or less elaborate network of individuals, whose task it was to realize the physician’s recommendations and the cook’s instructions.<sup>50</sup> The kitchen staff

<sup>49</sup> A similar phenomenon in the context of modern culinary culture can be seen in the continuing popularity of caviar, *foie gras* and champagne as status symbols, which is somewhat disproportionate to their culinary properties in comparison to other types of fish roe, poultry liver and sparkling wine.

<sup>50</sup> The household of Richard II employed a kitchen staff of 300 people in order to feed the 10 000 mouths that frequented it (Scully 1992: 84). On the other end of the spectrum, an urban bourgeois housewife cooking some of the simpler dishes for her family and a few guests would have been able to manage the task with the help of her daughters or a few servants (Redon, Sabban and Serventi 1998: 17).

of a great household was organized into various offices responsible for different aspects of its operation.<sup>51</sup> The fact that major medieval households usually kept a separate budget and separate accounts for the different offices meant that in addition to the manual labourers, there was also a need for lettered clerks, which provided career opportunities for men of humble birth who managed to acquire at least a modest degree of literacy (Keen 1990: 166–167).<sup>52</sup> The attested presence of these variably literate officials in the kitchen and its various departments is especially significant in terms of medieval culinary writing, since they may well be the missing link between the oral and practical knowledge of the professional cook and the written recipe collection.

### The physician

The medieval physician, whether as a retained member of a noble household or as an independent consultant, wielded considerable authority in determining the foods that were prepared for the lord of a household, his retainers and his guests (Siraisi 1990: 37; Scully 1992: 186). In addition to consulting with the steward and the head cook about the day's menu, a resident household physician would also sit at the master's table during the meal in order to ensure that he did not eat anything that would harm or disturb his health (Scully 1992: 42). Since every significant noble household in late medieval Europe employed at least one doctor who would by necessity be knowledgeable in current medical doctrine, "there can be little doubt that the cooks of these households were required to be guided directly or indirectly by this doctrine" (Scully 1988: 21).<sup>53</sup>

These medical practitioners, operating outside the academic milieu and, judging from the appearance of a large body of vernacular medical writing by the late 14<sup>th</sup> and early 15<sup>th</sup> centuries (Siraisi 1990: 20), often more comfortably literate in the vernacular than in Latin, occupy another key position with regard to the tradition of culinary writing and especially to its relationship with medical writing. According to Siraisi (1990: 32–34), this group of "ordinary medical and surgical practitioners" were eager consumers of any formal medical learning that they could get their hands on. The numerous "compendia, abbreviations and recipe collections"—mostly based on works produced in more "learned milieus" (32–4)—circulating among them also made them potential consumers of culinary recipe collections, which are often found among medical material. Despite the claim of Voigts (1989) that culinary texts "are not commonly met with in the company of

<sup>51</sup> The *pantry* made or bought and distributed the bread, and also encompassed the waferer and the laundresses. The *butlery* (or *buttery*) brewed or bought the ale and delivered the wine to the table, while acquiring and storing the wine was the responsibility of the *cellar*. The *kitchen* was supported by its subordinate offices in preparing and delivering the food: the *larder* was responsible for meat and fish, the *poultry* for birds, *scullery* for pots, pans, and fuel, the *saucery* for the sauces, the *pastry* for pie shells and other pastries, and the *spicery* for the spices it received directly from the great wardrobe (Hammond 1993: 122–3).

<sup>52</sup> For example at the turn of the 15<sup>th</sup> and 16<sup>th</sup> centuries, Cardinal Wolsey's kitchen had two clerks of the kitchen, a clerk-controller, a surveyor of the dresser and a clerk of the spicery (Strong 2002: 89).

<sup>53</sup> In addition to the physician, the medical profession was in larger households also represented by the pharmacist or apothecary. Being responsible for all drugs used in the household—including the spices and sugar used by the kitchen—he also played a significant role in the preparation of food. Before the cook could prepare a meal, he had to submit a requisition to the apothecary for any spices he intended to use (Scully 1992: 30).

‘scientific’ writings” (348), this cannot have been all that uncommon, as both of the versions of the *Potage Dyvers* collection that are found in miscellany manuscripts (MSS Ad and C) are accompanied by medical texts which can be considered to represent the more popular end of ‘scientific’ writing (see section 9.2 and appendix F for a description of the manuscript contexts of the different versions).<sup>54</sup> Weiss-Amer (1992: 78) finds further evidence for this connection in the form of culinary recipe collections owned by physicians, and points out that

in addition to aristocratic cooks and the circulation of culinary manuscripts, physicians trained in a particular cultural environment—such as thirteenth and fourteenth-century France and Italy—also contributed to the spread of culinary recipes and cooking practices.

(Weiss-Amer 1992: 78)

### The cook

The profession of a cook in general does not seem to have been overly prestigious or especially well paid in the Middle Ages. It was only through being employed by a wealthy bourgeois or noble master, that the cook could stand any chance of enjoying any kind of status (Scully 1992: 239).<sup>55</sup> Within the household, however, the head cook seems to have been a rather important figure. Since the cook quite literally held his master’s life in his hands, he had to be absolutely trustworthy, and this trust also translated into certain privileges, on par with the most select of the master’s aristocratic servants (Scully 1992: 252).<sup>56</sup> It was in the kitchen, however, where the cook’s authority was most visible, even if he was by no means a sovereign ruler of his kitchen (Scully 1992: 241).<sup>57</sup> The memoirs of Olivier de Marche, quoted by Scully (1992) paint a vivid picture of the authority and role of the head cook in the ducal household of Burgundy in the mid-15<sup>th</sup> century:

The Cook orders, regulates and is obeyed in his Kitchen; he should have a chair between the buffet and the fireplace to sit on and rest if necessary; the chair should be so placed that he can see and survey everything that is being done in the Kitchen; he should have in his hand a large wooden spoon which has a double function: one, to test pottages and brouets, and the other, to chase the children out of the

<sup>54</sup> Also the recipe collection in MS Harley 5401 edited by Hieatt (1996) is reported by her to occur in “a fifteenth-century volume otherwise entirely devoted to medical matters”, as “do a high proportion of medieval culinary collections” (54).

<sup>55</sup> An extreme example of this is provided by the career of the celebrated author of the 14<sup>th</sup> century *Viandier*, Guillaume Tirel (1310?–1395), better known as Taillevent. He first appears on the culinary scene as a kitchen boy in the household of Jeanne d’Evreux (the granddaughter of king Philip III) in 1326. By the middle of the century, he was in the service of king Philip VI and a few years later of the dauphin, the son of John II. In 1364 the dauphin ascended on the throne as Charles V, and by 1373 Taillevent was described as the “premier queux du roi”. The last mention of him in a professional capacity is from 1392, and when he died in 1395, he was buried in a splendid tomb in the priory of Notre Dame à Hennebont (Wheaton 1983: 20–21).

<sup>56</sup> The cook could, for instance serve his master a dish himself, taste it for him in his presence and then go drink from the master’s buffet (Scully 1992: 252).

<sup>57</sup> Despite all of his authority, the head cook had to work together as a team with the various other household officers in order to ensure the smooth and efficient running of the household as a whole (Scully 1992: 241).

Kitchen, to make them work, striking them if necessary.

(Scully 1992: 243-244)

It is interesting that although the preparation of food in medieval Europe seems to have been primarily the task of women (Bynum 1987: 190),<sup>58</sup> the culinary elite seems to have been dominated by men, and all culinary texts whose (alleged) author is known, are attributed to men.<sup>59</sup>

Unlike the university educated physician, the cook did not acquire his profession through formal education, but through a long and arduous apprenticeship under an established master.<sup>60</sup> This hands-on manner of education also served to promote the international flavour of medieval cuisine: after having served their apprenticeship in one region, a cook might well take up occupation in another, learning there a different tradition of favourites to which he would add his own specialities (Scully 1992: 203). Established cooks of great aristocrats also frequently accompanied their masters as they moved from region to region in the network of medieval European politics.<sup>61</sup> The full implications of this primarily oral, empirical and non-literary method of education with regard to the role of the culinary recipe collection are not yet completely understood,<sup>62</sup> but it seems clear that the medieval cook—unlike the modern amateur—did not rely on written recipes in preparing his dishes:

The professional cook in the Middle Ages worked primarily from an image that was firmly fixed in his mind by experience—an image we might qualify as being gustatory and tactile as well as visual; everything he did while he was cooking worked towards a realization of that image.

(Scully 1992: 220)

## The kitchen

Judging from surviving examples and literary descriptions, the kitchen of a late medieval castle or a manor house was usually a room or a series of rooms that pro-

<sup>58</sup> Bynum (1987: 190) points out that for example the best known medieval French cookery book, the *Viandier de Taillevent*, assumes basic cookery to be done by women. The Vatican version of the *Viandier* justifies the omission of instructions for preparing such basic items as cabbage, leeks and veal with saffron and pepper sauce by remarking that “femmes en sont maistresses et chascun le sçait faire” (“women are experts with these and everyone knows how to do them”) (Scully 1988: 217-8, 295).

<sup>59</sup> This would seem to indicate that the situation may not have been much different from that of our time: while the majority of everyday cooking is carried out by women, most of the celebrated chefs are still men.

<sup>60</sup> The lack of formal education should not be taken to indicate lack of professional ability, or even of theoretical knowledge. As Scully (1992: 253) points out, the medieval cook was not only a craftsman (and occasionally an artist), but also “a professional who understood the laws of theoretical physics that must govern all that went on in his kitchen”. If we accept the hypothesis that medieval recipe collections reflect actual practice instead of mere wishful thinking, they may well be able to provide a glimpse at the degree of this scientific knowledge exercised by the leading culinary practitioners of medieval Europe.

<sup>61</sup> This is witnessed by the discovery of a Catalan cook at an English court, a Savoyard cook following a peripatetic Duke who was later to become Pope, a German cook in the service of an actual Pope in the Vatican, and recipes written in Italian with a marked Catalan accent (Scully 1992: 206).

<sup>62</sup> Studies on the registers employed in culinary manuscripts—especially with regard to the oral-literary cline proposed by Leckie-Tarry (1995)—are one possible avenue of examining the traces left by oral practices in the written recipe collections.

vided enough space to allow the staff of five to fifty persons to work at once at their tables, sinks and mortars. In addition to the main room—or the kitchen proper—it could also include a separate bakery or pastry kitchen and storage facilities such as the pantry and the buttery (Scully 1992: 88).<sup>63</sup> Since the medieval kitchen was a source of noise, smells, heat and smoke, and posed a very real fire hazard, it was often located slightly apart from the living quarters of the house, perhaps connected to the hall by a covered walkway, with a serving area—or a *dresser*—at or outside the hallside end of the kitchen, where dishes were laid out for the waiters to be carried into the hall (Wood 1983: 247-52; Scully 1992: 86-7; Brears 2012: 192-4).<sup>64</sup> The kitchen itself, where most of the actual cooking took place, also needed to be well ventilated in order to provide oxygen for the fireplaces or open hearths, and to allow for sufficient airflow for carrying the produced smoke up chimneys or out of roof louvers. Cleanliness in general seems to have been a high priority at least in the larger medieval kitchens; according to Scully (1992), a good fifty percent of the kitchen staff could consist of scullions, “labouring in the unglamorous tasks related to cleaning” (87).

Although many large kitchens were equipped with ovens,<sup>65</sup> most cooking operations were performed by the fireplace or above an open hearth. The fireplaces of a medieval kitchen were frequently fitted with a variety of devices designed to regulate the amount of heat applied to the contents of a pot or pan or for holding the pieces of meat or fish to be cooked (Scully 1992: 93).<sup>66</sup> In addition to the cooking fires and the implements used for the actual cooking, the medieval kitchen was equipped with a wide variety of miscellaneous utensils, including “huge stirring spoons, knives, graters, rasps (for cleaning wooden surfaces), hooks, hampers, tripods and oven-shovels of various types and sizes” (99). Due to the predilection of medieval cuisine for fine textures, two pieces of equipment enjoyed especial prominence: the mortar and the sieve, both of which were used to reduce ingredients to small particles and smooth pastes (99). With this quite extensive selection of equipment—and substantial manpower at his command—the medieval

<sup>63</sup> Larger kitchens could also have other specialized rooms for the uses of the various offices, as well as a separate ‘dishing-out area’, where the finished dishes would be picked up by the servers and carried to the hall (Scully 1992: 88).

<sup>64</sup> For example the early 16<sup>th</sup>-century kitchens of Stirling Castle in Scotland consisted of a series of three rooms, of which the middle one held hearths and the other end room a baking oven (Fawcett 1999: 16). The medieval kitchen of the Kenilworth Castle in Warwickshire, on the other hand, consisted of a single large room, with the foundations of hearths and ovens still visible. The buttery and pantry were set on either side of the passage from the kitchen to the great hall and there was a serving area on the middle level of the Strong Tower (Renn 1991: 6). Both of these kitchens are also set apart from the Great Hall and against the main wall of the castle.

<sup>65</sup> The medieval oven was very simple and resembled the stone ovens of later periods: a stone cavern which was heated by a fire burning within it and the coals removed once the stones had become hot enough for baking (Scully 1992: 95).

<sup>66</sup> A bracket or trammel allowed a cooking pot to be suspended over the fire without the need for a tripod, and if its horizontal arm was mounted on a vertical pivot, it also enabled the cook to easily and quickly move the pot over the fire or off of it (Scully 1992: 93). Grills were in common use for roasting relatively flat items such as fish, either whole or filleted, and also for supporting flat-bottomed frying pans. For the roasting of meats and round fish, a variety of spits were used, the size of which was determined by the size of the item to be roasted. The spit was hung on clips on a pair of metal stands or on the andirons either over or beside the fire, and turned by a spit-turner, who was often a young boy (94–95). Metal baskets filled with coals or heat-retaining stones were also used to keep dishes warm while waiting to be prepared for serving (94).

cook turned his raw ingredients into finished dishes, designed to please not only the palate, but also the discerning eye of his noble masters.

### 7.2.2 From the great hall to the private chamber: the daily meal

The most important building of a medieval aristocratic house was the great hall, which was often located alongside the central court. It served as the physical and metaphorical centre of the household, acting as “an estate office, parish hall, court room, living-room, dining-room, bedroom for servants and thoroughfare for people passing into the inner chambers of the building” (Paston-Williams 1993: 68). Thus it also had an important social function as the “central point of the estate” (Wood 1983: 49) and served as the site of all public and ceremonial occasions, including formal dinners and feasts. Its physical proportions were thus suitably impressive; as a representative example, Paston-Williams (1993: 68) mentions the halls of Great Chalfield, Cotehele and Rufford manors in Lancashire, which all measured 12 meters in length, 6 meters in width and 9 meters in height. One end of the hall usually had a *dais*—a slightly elevated section of floor—and the other a wooden screen that created an entrance passage from the main door. Behind the screen, another passage could lead to the kitchens (Keen 1990: 163). As to permanent furniture, there was very little: perhaps a chair or two on the dais for the master of the household and a *buffet* or a sideboard. Since the great hall was the public centre of its master’s life and served as a public display of the family’s honour and dignity, it would often be decorated lavishly with paintings and tapestries, along with banners and shields emblazoned with the arms of the family; in other words, “the scenic decor for a demonstration of nobility” (Scully 1992: 166–169).

#### The medieval dinner table

When the great hall was used for dining, trestle tables and benches were set up along its length.<sup>67</sup> The head or *high table* was situated on the dais at the end of the hall, the result forming a square U-shape with the diners seated on the outside, leaving a serving space along the middle. The medieval table was covered with a clean linen tablecloth and sparsely set compared to a modern one. Before each dish left the kitchen, it would be portioned into *messes*, designed to be shared by two or four people (Paston-Williams 1993: 67). *Trenchers*—square slices cut from stale bread—were used as plates: diners took pieces of meat and other solid food from the shared serving bowl either with the tip of their knife or with the fingers of their right hand and placed them, along with dabs of sauce, on his trencher, while more liquid foods were eaten directly from an *écuelle*—a shallow serving bowl normally shared by two diners. Whole fish and birds, as well as rabbits, lambs and quarters of veal were presented to the high table and carved in the view of the diners (Paston-Williams 1993: 77).<sup>68</sup> Knives and spoons were usually

<sup>67</sup> According to artistic depictions of dining, the rectangular trestle table had overthrown the round table of earlier age by the beginning of the 12<sup>th</sup> century, and become very much a universal feature (Strong 2002: 60).

<sup>68</sup> This meant that the skill of carving was an important part of any young nobleman’s education and explains the existence of detailed manuals and special vocabulary on the topic (see Marttila 2009).



provided by the diners themselves, but could occasionally be provided by the host as an especially magnanimous display of affluence. Even drinking vessels were not usually set at the table but provided by the pages tending the buffet at request, filled with the beverage of choice (Wheaton 1983: 5; Hammond 1993: 77; Scully 1992: 170-1). Even when drinking vessels were set on the table, they were—like the servings of food, known as *messes*—intended to be shared by two, three or four people (Strong 2002: 109).

Down to the minutest practical details of the table, it is the idea of sharing that best characterizes the traditional form of medieval dining: not only was the master of the household sharing his food with his guests, but the guests themselves were expected to share amongst themselves everything that was put in front of them (Scully 1992: 172). This communal nature of dining, as well as many of the conventions of the table, had their roots in the monastic environment and the communal *refectory* of the monks (Strong 2002: 52-4).<sup>69</sup> In the late medieval period, however, these traditional habits were in the process of change, spearheaded by the very top levels of society.

From about the middle of the 14<sup>th</sup> century it became more common for royalty and great lords to eat in public only on certain occasions—such as significant feast days—and by the 15<sup>th</sup> century this practice had become the norm (Mennell 1986: 88; Keen 1990: 164), although the old habit of the entire household dining together seems to have survived longer among the shire knights and gentlemen than among the urban upper classes (Strong 2002: 93-94).<sup>70</sup> In both England and France this correlated with the multiplication of private rooms in the manor-houses and castles, which afforded a greater degree of privacy to its owner and his family (Wood 1983: 61, 67).<sup>71</sup> In the 15<sup>th</sup> century, also very important guests came to be entertained to meals in the lord's private chambers, while those of lesser estate would sit in the hall, presided over by the steward of the household (Paston-Williams 1993: 73). The same sense of increased social division that is apparent in this new habit of dining in one's private quarters also manifests itself on the remaining public occasions. Even at the end of a more public meal or a feast (discussed in more detail below), the host could select a few privileged guests to accompany him to the privacy of his chamber for confections (*dragees*) and entertainment (Strong 2002: 113).

<sup>69</sup> In the 13<sup>th</sup> century, the importance of communal dining for the maintenance of harmony and order in the household was explicitly recognized by Robert Grosseteste (ca. 1170-1253), who emphasized its role when instructing the Countess of Lincoln in the proper conduct of her household (Strong 2002: 67).

<sup>70</sup> The decline in communal dining was lamented by some social conservatives, such as William Langland (ca. 1330-1387) in his *Piers Plowman*, edited by Kane and Donaldson: "Elenge is þe halle, ech day in the wike, / Ther þe lord ne þe lady likeþ no3t to sitte. / Now haþ ech riche a rule to eten by hymselfe / In a pryuee parlour for pouere mennes sake, / Or in a chambre wiþ a chymenee, and leue þe chief halle / That was maad for meles, men to eten Inne" (1975: 412).

<sup>71</sup> When dining in the confines of the private room, at least in winter, the table seems to have often been erected in front of the fireplace for comfortable warmth, a screen protecting the diners from direct heat, as depicted in the January scene of Duc de Berry's *Les Très Riches Heures* (Longnon and Cazelles 1993: 2, 173).

### The organization of daily meals

In addition to differences in the foods served and eaten in the late Middle Ages, also the organization of meals differed from our modern day habits. Both meal times and the internal composition of the meals were largely governed by the dietic rules discussed above, which were intended to ensure that the nourishment provided by the food would be assimilated into the body as efficiently and safely as possible (Scully 1992: 118). As a rule, only two meals were eaten in a day, since medieval physicians considered it a most dangerous practice to eat before the previous meal had made its way out of the stomach and one felt hungry again (Scully 1992: 119). The first meal of the day, dinner or *prandium*, was normally the major one and was taken around noon, when half of the day's work was already done. The day was concluded with a second meal, *cena*, served some six hours later (Paston-Williams 1993: 63). Because it originally consisted of bread soaked in hot wine or broth, known as sops, it came to be called *supper* in English (Scully 1992: 118–119). During the 15<sup>th</sup> century, however, supper often came to be served later, at 7 or 8 o'clock, and was elaborated into a more complicated affair. Also *breakfast* began to appear, and by the late 15<sup>th</sup> century it seems to have been quite commonly taken.<sup>72</sup> According to Scully (1992: 120), the earliest breakfast “was undoubtedly just a chunk of bread and a mug of watered wine”, but we also have evidence of anchovies and other fillets of fish, probably in smoked form, being consumed as breakfast (Scully 1992: 120),<sup>73</sup> and by the 16<sup>th</sup> century, it would become “a substantial meal for those who could make it so” (Paston-Williams 1993: 63)

Also the makeup and organization of individual meals was governed by a framework of regulations and conventions of several kinds. The cook planning the menu had to know whether the day or the season prohibited red meats and what foodstuffs would be available to him at that time of the year. He would have received instructions from his master or the steward about the number of courses to which the various dishes would have to be grouped. He would have to know the function of each dish in the meal and whether a *subtlety*—a spectacle food functioning as a showpiece—was required. And finally, he would have to be able to prepare dishes that were not harmful to the temperament of his master or his guests, taking into account the current season and the climate (Scully 1992: 110). Although these considerations had their impact on every meal of an aristocratic household, it was the medieval feast, in which they—as well as many other aspects of medieval dining—found their quintessential expression.

### 7.2.3 A special occasion: the medieval feast

[T]he characteristic medieval meal was the feast, and it was more an aesthetic and social event than a gastronomic one. The feast was a

<sup>72</sup> The accounts of Dame Alice de Bryene at the beginning of the 15<sup>th</sup> century regularly allowed for it, and the household ordinance of Edward IV from 1478 specifies that residents of his court down to the rank of squire should have breakfast. It is also known that the son of Edward IV breakfasted after hearing morning mass (Hammond 1993: 105; Strong 2002: 104–5).

<sup>73</sup> Not needing to be cooked or otherwise prepared, they would be a convenient means of holding off hunger in the morning.

banquet for all the senses; indeed, food was almost an excuse for indulging senses other than taste. Medieval chroniclers who describe feasts do not give menus, although they lavish attention on the entertainment provided. They describe the appearance of dishes, not the flavor; the sequence of events, not of courses. Medieval cookbooks make it clear that visual effects were more important to a medieval diner than taste and that vivid colors [...] were often applied at the expense of flavor. (Bynum 1987: 60)

Although most food historians do not share Bynum's view of the medieval cooks' and diners' disregard for taste—shared also by some other traditional historians like Braudel (1981)—it is true that the medieval feast was much more than just a meal. According to Goody (1982: 141-2), there existed a conceptual contrast in medieval England between feasts and ordinary meals, and they can be seen to have been atypical of the general eating habits in two respects: not only did they involve only a small minority of the population, but even for them, they were “high points of an oscillating dietary regime” (Mennell 1986: 22). Although banquets and feasts were exceptional as social events and marked a deviation from everyday activity, Scully (1992: 123) points out that from the culinary standpoint, the dishes served during a grand banquet were very much of the same kind that was served during ordinary meals in a similar household, the difference lying mainly in quantity and in the presence of the spectacular *subtleties*.<sup>74</sup> Looking at the contents of English cookery collections, such as the *Potage Dyvers*, it does indeed seem likely that some of the dishes would not have been served on an everyday basis even in the royal household. On the other hand, the vast majority of dishes contain nothing that would make their everyday preparation in a wealthy household unlikely.

### Symbolism and ceremony

In the early and high Middle Ages up to the 12<sup>th</sup> century, feasts had an important social and political function. They were not merely occasions for entertainment and celebration, but a way for the nobility to assert their social rank and power (Mennell 1986: 58). The feast was

a quintessential part of the fabric of feudalism, a massive periodic culinary event celebrating the relationship between a lord and his vassals, and the power this relationship engendered. (Strong 2002: 66)

At the turn of the 13<sup>th</sup> century, however, there seems to have occurred a shift in the atmosphere surrounding the feast, as the new ideals of *courtliness* began to shift the feast from a ritual of feudal dependence into a manifestation of *friendship* (Strong 2002: 66). Whatever the specific social virtue they embodied, medieval feasts could be described as “realizations of aesthetic and social ideals” (Wheaton

<sup>74</sup> Grieco (1992: 37), however, presents contrary material from Italy, coming to the conclusion that the dishes described by Italian cookbooks were not in fact everyday food but dishes served only on special occasions.

1983: 1).<sup>75</sup> Despite these lofty ideals, the role of the feast as a demonstration of the wealth and power of the host and of the prevailing social hierarchy never disappeared (Wheaton 1983: 3):

In theory, a feast was the epitome of love and fellowship, and while reality often failed to mirror this ideal, a ceremonial dinner was a visible demonstration of the ties of power, dependence and mutual obligation which bound the host and guests. It was politic for the host to appear generous, because the lavishness of his table gave a clue to his resources; it was wise to be both hospitable to dependents and discriminating in the choice of guests of honour, because the number and calibre of diners in the hall revealed his importance and his power.  
(Henisch 1978: 56-57)

Everything from the seating order to the trappings of the feast hall served to demonstrate the host's standing and his relationship to his guests. Everyone present at the feast was seated according to their order of precedence, with the highest ranking people sitting at the *high table* with the lord, and the humblest at the far end of the hall. The table to the right of the lord was the second in dignity and was known as the *Rewarde*, because it was served with the dishes from the lord's own table, followed by the table opposite it and so on, progressing down the length of the hall (Paston-Williams 1993: 71). The high table was always served first, with the food subsequently carried to the table next in rank and so on, until it ran out (Scully and Scully 1995: 44). This meant that even in more lavish households, the first, simpler or more substantial course was all that many members of the household would receive, the more delicate and refined dishes being reserved for the higher ranks and honoured guests (Hieatt and Butler 1985: 5; Hammond 1993: 118; Paston-Williams 1993: 65).<sup>76</sup> This hierarchical distribution of food was sometimes noted also in recipes; for example a recipe for pike and eel in broth found in *The Ordinance of Pottage* (edited by Hieatt (1988)) instructs to "serve hole pykys for lordys quarters for othir men" (40), while a recipe for conger, turbot, and halibut suggests that one should "serve congure ii or iii pecys on a chargeor for thy soveraynys, [...] serve the remnaunt for othir men" (103).<sup>77</sup>

Also the extremely formulaic etiquette observed at a feast served to emphasize the nobility of the host and his guests and to set them apart from common society (Scully and Scully 1995: 41). Even the physical setup of the feast hall with its U-shaped tables contributed to this effect: the arrangement not only facilitated service from the inner area, but also promoted formal and ceremonious presenta-

<sup>75</sup> The role of the feast as an ideal image of society was heightened by the increasingly elaborate ceremonial enveloping it. In the chivalrous romances of the Late Middle Ages, the feast figures as a "symbol of joy and harmony, an occasion for the display of the virtues stemming from good breeding and the exercise of courtesy" (Strong 2002: 101).

<sup>76</sup> For a concrete example of this practice we can turn to the coronation feast of Richard III in 1483, where only the king's table were to receive the full 3 courses served, while the lords and ladies had to settle for 2 and the commoners for only one, presumably the one containing only the humblest dishes. The 'lesser delicacies' were also served to the lords and ladies, but for instance peacock (see recipe PD 128 in the present edition) was reserved exclusively for the king (Hammond 1993: 135; Strong 2002: 104; Lehmann 2003: 27).

<sup>77</sup> The recipes in *Pottage Dyvers* do not contain such status-bound serving instructions, although they do on several occasions specify the number of items to be served in a single vessel.

tion of dishes and afforded a focal point for the *entremets* and any other entertainment that the host wished to offer his guests as a display of his taste (Scully 1992: 169–170). Also the buffet (or *dressoir*) that had originally been merely a serving aid, had by the 15<sup>th</sup> century assumed a life of its own, being dedicated solely for the display of the most valuable household plate of gold, silver or pewter (Scully 1992: 166–8; Strong 2002: 96). A similar function was played by the decorative *salt cellar*, which was laid on the high table at the right hand of the lord and would have been the most magnificent piece of silverware that the lord could afford.<sup>78</sup>

### The course of a feast

The ritual for grand dining seems to have been the same all over late-medieval Europe, with only minor variations in its organization. Every meal began with a hand-washing, often with scented water.<sup>79</sup> After this the first course would be carried in, the accompanying ceremonial being determined by the grandness of the household (Strong 2002: 105–106). At a medieval feast, dishes were not served individually but in *servings* or *courses*, consisting of a variety of dishes, depending on the affluence of the household. The usual number of courses seems to have varied geographically, from the two to three courses in 15<sup>th</sup>-century England or Savoy to the grand and formal dinners of 8–12 courses in Italy (Hammond 1993: 131; Scully 1992: 132), with French menus indicating a number of courses somewhere between these.<sup>80</sup> The number of courses and dishes was also influenced by the scale of the event and the place of the diner at the table, determined by his status: for example the coronation feast of Richard III in 1483 consisted of a total of three courses, of 15, 16, and 17 dishes, respectively (Hammond 1993: 134), of which only the king's table received all three. The bills of fare included in the *Potage Dyvers* manuscripts indicate meals of one to four courses (the majority having three) of 6 to 22 dishes (with a median of 12),<sup>81</sup> representing occasions of varying size.<sup>82</sup>

The dishes of a course were normally placed simultaneously on the table, each guest then selecting what they wanted from the dishes placed within their reach (Redon, Sabban and Serventi 1998: 10; Strong 2002: 109). Although the logic of the sequence of dishes in surviving menus for medieval feasts may not be apparent at a first glance, the ordering of dishes into servings was not random. Instead, it was—in theory, at least—based on the established idea of the stomach as a cooking pot and the relative ease of digestion of the various foodstuffs (see subsection 7.1.1 above) (Scully 1992: 128–29; Strong 2002: 109). While the same basic principles seems to have been followed in structuring the feast in both England

<sup>78</sup> The salt cellar could take various shapes, the unifying feature being its precious material. According to his will from 1380, Edmund Earl of March had a silver salt cellar in the shape of a dog, while the will of John of Gaunt mentions a gold one embellished with a garter. Salt cellars in the shape of a ship (known as the *nef*) were especially common in France (Hammond 1993: 109; Scully 1992: 171).

<sup>79</sup> For monarchs and higher nobles (down to the rank of earl in England), this was followed by the rite of *assay*—testing everything for traces of poison, using either a unicorn horn (a narwhal tusk) or a serpent's tongue (a fossilized shark's tooth) (Strong 2002: 105–106).

<sup>80</sup> The scale of some 15<sup>th</sup>-century Italian *banchetti* organized in Venice during the Carnival was apparently so extravagant that the Senate of Venice protested against them and in 1460 forbade banquets costing over half a ducat per diner (Braudel 1981: 188), apparently in vain.

<sup>81</sup> The largest courses seem to consist of a large number of different species of birds or fish.

<sup>82</sup> A description of all the bills of fare included in each of the manuscripts is included in appendix F, and a reading edition of their contents is found in appendix C.

and France, there seem to have been some differences as well, noted by Flandrin (1992). The French menus seem to indicate a more elaborate organization, with more courses and different types of foods associated with each course, the progression from more substantial foods to more delicate ones taking place over the whole meal, while the English menus seem to repeat the progression also *within* each course, essentially structuring them as a series of smaller meals, with each course repeating the same pattern and also becoming more digestible as a whole than the previous one (Lehmann 2003: 28).<sup>83</sup> With these differences in mind, let us then examine the order in which a medieval feast at its most elaborate would have progressed.

The very first item of the meal was usually something that could be called an *aperitif*, from Latin *aperire*. Its purpose was to open up and warm the stomach to receive the food to be digested.<sup>84</sup> Without an aperitif, there was a danger that the stomach would start its work sluggishly and some of the food to follow would remain undigested (Scully 1992: 130). Next, the stomach was ready for the first substantial courses, consisting of a moderately warm and moist *potage* or *brouet* or some other boiled dish.<sup>85</sup> After this were usually served one or more courses of roast meat or fish, together with their appropriate sauces (Scully 1992: 128; Redon, Sabban and Serventi 1998: 11).

After the serving of roasts (or between them, if several courses of them were served), there could be an intermission with its *entremets* or *subtleties*—spectacles of food or other entertainment.<sup>86</sup> As mentioned above, in English menus this progression of pottages followed by roasts and terminated by a *subtlety*—the English parallel of the *entremet*—would seem to occur within each course.<sup>87</sup> The *subtlety* (or *sotelte*) was essentially a table decoration, which could be an ornament made

<sup>83</sup> This repetitive structure noted by Flandrin does not seem to occur in English menus for smaller feasts, and it may well be a feature of the grand feasts held for the royalty (Lehmann 2003: 28).

<sup>84</sup> The aperitif could take the form of either solid food or a liquid. There was no widespread agreement on the best aperitif, but many physicians recommended confections consisting of certain seeds steeped in honey or sugar. Anise, caraway, fennel and cumin were held to be the warmest of seeds and therefore especially suitable for stoking the digestive fire of the stomach. Also wine—especially red wine—when drunk in a moderate quantity on an empty stomach was held to open its orifice and arouse the appetite (Scully 1992: 129–130). Foods containing acids were apparently also considered to open the stomach and therefore be suitable aperitifs (Redon, Sabban and Serventi 1998: 11).

<sup>85</sup> There seem to have been two opposing interpretations of the digestive qualities of these types of foods, depending on the view taken with respect to the debate mentioned above: they were either considered to require a long time to cook due to their moistness, or to digest quickly due to their closeness to the temperament of the human body. In either case, these types of dishes were recommended to be served first and seem to appear in this position in surviving menus (Scully 1992: 128; Redon, Sabban and Serventi 1998: 11).

<sup>86</sup> According to Strong (2002), “it is clear that by 1400 the word *entremet* referred to various manifestations appearing in the intermissions between courses at great banquets” (118). It could be anything from a single dish—usually somehow spectacular—to a collection of such dishes, or even a grandiose spectacle produced primarily by carpenters and artists instead of the cook. In late 14<sup>th</sup>-century France *entremets* sometimes took the form of theatrical performances, staging the events of romances and legends in the hall. This idea quickly spread to Burgundy, Savoy and Spain, but does not seem to have reached England until the early 16<sup>th</sup> century (Scully 1992: 108–10; Strong 2002: 123–4).

<sup>87</sup> This makes it somewhat uncertain whether all of the items in the course would be placed at the table simultaneously, or—which seems more likely from a logistical point of view and is also suggested by Flandrin (1992)—brought in as a progression of dishes, each being served in turn to the high table, and then passed on down the hall, essentially making each course a stream of dishes following each other.

entirely of sugar or marzipan, or a composite display piece which was only partially edible. The subtlety often contained allegorical references, and its object was to impress the guests with the skill of the cook (or other associated parties) and thus with the equal cleverness of the host in employing him (Hammond 1993: 142; Strong 2002: 120). This should not be taken to mean that the *entremet* or *subtlety* was the only form of entertainment at a feast: music and song seem to have been equally popular and frequent features of medieval feasts. For example at the knighting of the eldest son of Edward I in 1306, guests brought their own minstrels to sing *chansons de geste*, and Edward I himself had 27 minstrels in attendance. The minstrels would entertain the guests by singing and playing various instruments, including harp, psaltery and lute, along with a variety of wind and percussion instruments (Strong 2002: 123).

After the intermission, the dinner could resume with a *desserte*—consisting of more delicate dishes<sup>88</sup>—and finally concluded with what was in France called the *issue de table*—a course of cheeses,<sup>89</sup> candied fruits and cakes served with sweet wine—whose purpose was to serve as a digestive and to close the aperture of the stomach.<sup>90</sup> A similar function was served by the *boute-hors* or *voide* that was served in another room after the meal proper and consisted of *dragées*—sugar candies spiced with warm spices such as ginger and cinnamon—and candied seeds of a warm nature, again served with sweet spiced wine (Scully 1992: 130-1; Redon, Sabban and Serventi 1998: 11). The feast formally concluded with the lord and the guests seated at the high table washing their hands again and the lord standing up to drink a toast signalling the end of dinner (Paston-Williams 1993: 79). Taken as a whole, the late medieval feast—like the modern Christmas dinner—was thus an extensive affair regardless of its social scale, serving as an opportunity for the household to display its very best in every regard.

### 7.3 Conclusion

While it is naturally impossible to bridge the gap between the cultural and situational context of late Middle Ages and the present day, this chapter has hopefully provided the modern reader with an awareness of the cultural and ideological constructions on which the internal logic of the *Potage Dyvers* and other Middle English recipe collections was based, as well as the social and physical context in which the food described in them was prepared, served and consumed. As was observed in chapter 6, the tradition of medieval recipe collections emerged at the same time as both the written word and the vernacular English language were asserting themselves as indispensable facets of everyday life. The strengthening of the socio-political position of the middle classes created a competitive situation

<sup>88</sup> These often included such sweet dishes as fruit in honey or sugar syrup, sweet fruit purees, sweet custards, fritters, crêpes and sweet tarts and pasties, but also small birds and other delicate meats belonged to this category (Hieatt and Butler 1985: 3-4; Scully 1992: 136; Redon, Sabban and Serventi 1998: 11).

<sup>89</sup> As was pointed out in *The human body and its functions* in subsection 7.1.1, cheese was considered to be heavy and to descend in the stomach, pressing the previously consumed foods to the very bottom, considered its 'hottest' part.

<sup>90</sup> This action was seen as analogous to fixing a cover on a cooking pot and based on the everyday observation that the cooking process occurs most efficiently in a closed container.

between them and the old aristocracy, where the traditionally literate aristocratic, learned and ecclesiastical powers sought to maintain their distinct identity and status, while the middle classes tried to emulate the upper classes and become a part of it. The spread of utilitarian literacy down the social scale also meant that various kinds of written works, including culinary recipes and other household guides and the practical information conveyed by them played an important role in this battle. Culinary recipes, such as the ones edited here also embody in a very concrete form several of the ideational frameworks that shaped the medieval world view, namely the view of the human body and its functions inherited from antiquity, the Christian doctrine about the relationship of the human soul to the material world, and the established social order and the importance of understanding and enacting one's place in it. Furthermore, the recipes are intimately connected to the physical conditions under which they were prepared and consumed, and by the limitations and possibilities offered by the medieval household and its kitchen.

Despite the increase in literacy and the general rise in standards of living after the great epidemics of the 14<sup>th</sup> century, the late medieval world view, including the relationship of man to God, the macrocosm, and fellow man, was still very much based on ancient and biblical authorities. Like all other aspects of life, these views also influenced the medieval culinary culture. The human body was seen to be a microcosm of the universe (Burrow 1988: 12-3), consisting of the same four elements, and nutrition was seen as one of the principal means of preserving their balance and maintaining health. This meant that the humoral properties of foodstuffs, cooking methods, and the resulting combinations were considered just as important as their culinary properties, which had a significant influence on the composition of both individual dishes and medieval meals as a whole. This system of humoral balance was complemented by the view of digestion as a form of cooking and of the stomach as a pot, which obligated the design of meals to take into account not only the humoral qualities, but also the varying digestive properties—or 'cooking times'—of different dishes.

In addition to the medical views about nutrition, late medieval food culture was also significantly influenced by the traditions and doctrines of the Christian church, which also had its specific views about the influence of food on the soul of man, partially based on the ancient medical theories. Since food—especially meat and animal products—were equated with the carnal body and considered to incite also sexual lust, its consumption had to be regulated. While the example of the eucharist legitimized the concept of communal eating, the institution of fasting, inherited from Judaism, attempted to limit the consumption of spiritually harmful foods by restricting the times when they can be eaten. This had an enormous influence on medieval culinary culture by forcing cooks to develop an alternative culinary tradition based on allowed ingredients—mainly a variety of fish—that was gastronomically equal to the meat-based dishes they were used to substitute, thus significantly increasing the prominence of these allowed foodstuffs in medieval cuisine.

As a highly social activity, eating was also subject to the hierarchical rules of medieval society. Not only did the position of a household in the social hierarchy—which was heavily correlated with its economic means—limit the material means available for culinary purposes, but it also determined the kinds of food and asso-



ciated ceremony that were expected of it. Similarly, within the household and its dining table, members and guests alike had their own places, determined by the prevailing social hierarchy, which also determined—in a very real sense—what they got to eat. As the god-given hierarchical organization of society began to unravel with the rise of the middle classes in the 15<sup>th</sup> century, eating and the social mores associated with it became an important symbolic weapon used by the aristocracy to distinguish themselves from the middle class, and a sought-after status symbol for the upwardly-mobile members of the middle class.

Far from residing merely in the sphere of culture or the written page, recipes are an eminently practical form of discourse, being intimately connected with the everyday running of a medieval household and its practices of food production and consumption. The dishes described by medieval recipes were—most likely—created and prepared by actual people, whose background experience and professional skill, as well as the physical conditions in which they practised their craft, set the practical limits to what could or could not be served at the table. Together with the cultural thought-styles described above the more or less routine practices of running the household formed the situational framework in which the food was served—dictating the times of everyday meals and the organization of celebratory feasts through which the master of the household expressed his status.

While the foodstuffs and procedures employed by medieval cooking are essentially the same as those of today, these differences in the context which determined the ways in which those foodstuffs and methods were combined to produce cooked food, resulted in a cuisine which was in many respects alien to observers of later times. This perceived alienness of medieval food has over the last three centuries provoked reactions mainly in the negative. The first modern editors of medieval recipes considered the dishes described by them to “produce an effect very unpleasant to a palate of this day” because of the strange mixtures of ingredients, and to be so spicy as to “be relished only by those accustomed to the high-seasoned dishes of the East and West-Indies” (Warner 1791: xxxiii). A hundred years later, Austin (1888) commented upon the very same recipes included in the present edition that they “would astonish a modern Cook”, and concludes that medieval people evidently had “stronger stomachs, fortified by outdoor life” (viii) to be able to eat such spiced food. These opinions were assumed also by cultural historians, who either disparaged medieval cuisine with gusto or ignored it entirely. For example the French historian Fernand Braudel (1981) sees “great French cooking” to have developed only in the 18<sup>th</sup> century (189) and considers medieval food to have been practically inedible, advising anyone against trying to actually prepare the dishes because “[a]ll experiments have turned out badly” (190). According to him, there existed no sophisticated cookery in Europe before the 15<sup>th</sup> century, and that the medieval feast was merely “an orgy of greed” where “quantity prevailed over quality” (190).<sup>91</sup>

Fortunately, the appreciation that modern globalization has given us for more ‘exotic’ contemporary cultures and their cuisines makes it much easier to appre-

<sup>91</sup> Braudel’s rather poor estimation of medieval cookery seems to stem, at least partially, from his reliance on the descriptions of 19<sup>th</sup>-century historians, rife with misunderstandings, such as the idea that a course in a medieval feast (according to him, called “*mets*” or “*assiette*” in French) would have consisted of a single huge vessel on which all of the meats, vegetables and fish would have been piled in a great pyramid, forming a “dreadful hotch-potch” (Braudel 1981: 190).

ciate also the cultural context and the associated cuisine of the European Middle Ages and to realize that it is in fact no more alien than, say, Indian or Chinese cuisine—which also have their own medical, religious and social traditions on which their foodways are based. In order to avoid the rash judgements of our predecessors, it is important that we approach and interpret the evidence provided by medieval culinary writing in terms of its original context instead of judging it based on our modern conceptions of the physiological and cultural functions of food.

## Chapter 8

# Medieval culinary writing

What a society eats is intimately related to what the society is and does. That food is one of man's absolute necessities means that it must be a principal object of study by anyone who seriously hopes to understand the history of humanity. (Scully 1988: 30)

While food is at its core is primarily a physical concern, its intimate and bodily nature also makes the production and consumption of food an important social concern, as was observed in chapter 7. While the physical necessity of food to the human body means that its roots are embedded in the tradition of *medical writing*—as will be shown below in the discussion of the history of the European tradition of culinary writing—the heightened importance of the social aspects of food in the late Middle Ages meant that culinary writing also became to be seen as writing about *social conduct*. The relative plenty of food in the 14<sup>th</sup> and 15<sup>th</sup> centuries meant that its role as “a consumer good able to create and disseminate cultural meanings” (Sponsler 2001: 13) became emphasized over its role as a subsistence necessity.

According to Hieatt (1998b), much of what was written about medieval food before the 1980s “was just plain wrong, whether because the evidence was misinterpreted or because it was still insufficient in quantity” (101). Until the wave of new scholarly editions of medieval culinary recipes in the 1980s, the general impression of medieval culinary culture was based on the few 18<sup>th</sup>- and 19<sup>th</sup>-century editions, of which the most influential was that of Warner (1791), who viewed the foods described by the *Forme of Cury* (FC) collection that he edited rather disdainfully as “French-inspired and thus disguised and metamorphosed into ‘complex and non-descript gallimaufries’” (Lehmann 2003: 19). While this characterization “tells us more about the eighteenth-century approach to French cuisine than about medieval practice” (Lehmann 2003: 19), it nevertheless influenced historian's judgements about medieval cookery until quite recently.<sup>1</sup>

During the last three decades, however, there has been increasing interest in medieval cuisine, and the field is at long last being taken seriously by historians

<sup>1</sup> See for example the descriptions of Braudel (1981), mentioned at the end of the preceding chapter.

“and even some philologists” (Heatt 1998a: 139–40), as this thesis and the present edition also testify. Grieco (1992: 29) has also seen the increase in the number of new editions of medieval culinary texts—not only English but also continental ones—as an indication of their increasing importance as sources for not only culinary history but social history in general. Even more than editions, the recent decades have seen a great increase in books and articles on the various aspects of culinary history, to the extent that it is impossible to present any kind of comprehensive review of it within the confines of this thesis.<sup>2</sup> Unfortunately for the textual scholar, the focus of most of these works is predominantly historical, and culinary manuscripts as such receive mostly passing remarks, although at least Scully (1992: 4–9), Redon, Sabban and Serventi (1998: 1–3) and Strong (2002: 78–87) do dedicate some pages to the issue of medieval cookery manuscripts.

In order to acquaint the reader with the tradition and characteristics of Middle English culinary writing, this chapter will first outline the wider European tradition before focusing on the surviving English recipe collections, the earlier editorial work done on them, and their status as evidence of actual historical practices. This overview of the material will be followed by a discussion of the textual properties of the *recipe collection* as a composite text or a *discourse colony*, including the implications this has on its internal structure and organization, and its propensity to itself occur as a component of larger discourse colonies like miscellanies and commonplace books. After looking at the recipe collection as a whole, I will then focus on the characteristics of the individual *recipe* as a text type and a genre based on earlier research, arguing that culinary recipes actually constitute a specific *register* of the wider genre of *recipes* or practical instructional texts. After examining the textual, linguistic and functional features of recipes, including their characteristic textual organization and the perceived vagueness of their contents, the final part of the chapter will focus on what we currently know about the production and use of medieval culinary recipes and of the people who were most likely to be involved in these processes.

## 8.1 The European tradition of culinary writing

The tradition of European culinary writing has long roots that go back at least as far as the Roman Antiquity. The status of the first known culinary treatise in the west is disputed, but a strong candidate would be the *Deipnosophists* of Athanaeus, a native of the Egyptian town of Naucratis, which was compiled around A.D. 200 and is a compilation of “the manners and the customs of the ancients”, allegedly gathered by the author from the writings of 800 different writers (Goody 1982: 103). The earliest surviving cookbook from the northern side of the Mediterranean is the *De Re Coquinaria*, attributed to the Roman gastronome Marcus Gavius Api-

<sup>2</sup> A good overview of the literature on medieval food and ‘foodways’ can be acquired by perusing the bibliographies of some of the culinary historical works cited in this chapter; especially Hammond (1993) and Weiss Adamson (2004) are useful starting points in this regard. For a brief overview of the characteristics medieval food, the reader is advised to consult the aforementioned works and Scully (1992), while those desiring a more detailed view of the different practical aspects of medieval English cooking, illustrated by modern redactions of medieval recipes, are pointed towards the award-winning recent work by Brears (2012), and the articles in Woolgar, Serjeantson and Waldron (2006) that provide a comprehensive view of the material history of medieval diet and nutrition.

cius who lived in Rome during the reigns of Augustus and Tiberius, but most probably written sometime during the 4<sup>th</sup> century, 300 years after his time. The oldest manuscript survivals of this work are from the 9<sup>th</sup> century, the most famous being the Codex Vaticanus Urbinas Latinus 1146, which also seems to be the ancestor of all the surviving 15<sup>th</sup>-century manuscript copies (Elo, Laaksonen and Valjakka 2002: 9–11).<sup>3</sup>

Another significant influence to the medieval tradition of culinary writing came from the Arabic cookbooks and health manuals that reached Europe both through the Crusader states and the Iberian peninsula.<sup>4</sup> According to Goody (1982: 129), these medieval Arabic cookbooks—much like their European counterparts are believed to do—described the cookery of the courts, which had developed “patterns of conspicuous consumption, based upon their Roman, Greek and Persian predecessors”. He also posits that they were not composed by cooks but by “great personages” who concentrated on their favourite recipes, leaving out more ordinary dishes. Also the conclusions of Rodinson (2001b: 130) and the frequent attribution of Arabic cookery manuals to princes, poets or historians instead of famous cooks, would seem to support this hypothesis.<sup>5</sup> Although they seem to have originated in the courts, it is interesting to note that as the Abbasid civilization came to an end, there was “a change in the numbers and level of cookbooks, which now fell into the hands of the academic and merchant middle class” (Goody 1982: 131).<sup>6</sup>

Although the first proper *cookery books*—which Strong (2002: 49) sees as the first evidence of the emergence of a more sophisticated cuisine—survive only from the 13<sup>th</sup> century, the first texts containing culinary recipes survive already from the 9<sup>th</sup> century. These precursors of the culinary collection proper are medical treatises belonging to the *regimen sanitatis* tradition, of which “[c]ulinary recipes and cooking instructions” form a part “from the earliest Pre-Arabist text to the end of the Middle Ages” (Weiss Adamson 1995: 204). When *regimina* began to be produced in western Europe in the 13<sup>th</sup> century—from Arabic and classical

<sup>3</sup> Although Mennell (1986: 52) sees a close and even “derivative” relationship between the cookery of ancient Rome and the medieval upper classes, he notes that there are also significant signs of discontinuity. Based on these discontinuities, he considers it likely that what continuity there existed rested mainly on the remnants of an oral tradition rather than any knowledge of Roman culinary texts such as *De Re Coquinaria*.

<sup>4</sup> Of these early (pre-1258) Arabic cookbooks, three have survived to our day. The oldest of them, the *Kitāb al-ṭabīkh wa-islāḥ aghdhiya al-ma'kūlat wa ṭayyib al-aṭ'ima al-maṣnū'āt* of Abū Muḥammad al-Muẓaffar Ibn Sayyār al-Warrāq—preserved in two copies, Bodleian Library MS Huntington 187 at Oxford and Helsinki University Library MS Arab. 27—has been dated to the second half of the 10<sup>th</sup> century and has a significant “dietetic leaning” (Goody 1982: 129; Rodinson 2001b: 102; Öhrnberg and Mroueh 1987: iii-iv). The *Kitāb al-wuṣṣā ila al-ḥabīb fi waṣf al-ṭayyibāt wa al-ṭīb*, which was probably written around 1260, seems to have been the most widely circulated of the three, since it survives in a total of ten manuscripts (Rodinson 2001b: 116, 126-9). Finally, we have the *Kitāb al-ṭabīkh* of Shams al-dīn Muḥammad b. al-Ḥasan al-Baghdādī, which was written in 1226 and has been both edited and translated into English (Rodinson 2001b: 102; Öhrnberg and Mroueh 1987: iii-iv). These collections, however, do not seem to be the first that were written; it is very likely that the first Arabic cookery books appeared in the second half of the 8<sup>th</sup> century (Öhrnberg and Mroueh 1987: v).

<sup>5</sup> Despite this, these culinary recipe collections seem to have been practical rather than literary works, as Rodinson (1949) points out in describing the *Kitāb al-ṭabīkh* of al-Baghdādī: “C'est un livre de praticien, les recettes sont claires, précises, sans aucune trace de littérature” (104).

<sup>6</sup> Although there occurred no such upheaval in late medieval Europe, a similar process seems to have been underway there, in connection with the rise of the middle classes.

sources—the number of culinary recipes contained in them rose sharply, shifting some later *regimina* closer to actual cookbooks than medical treatises.<sup>7</sup>

According to Weiss Adamson (1995: 201–202), these *regimina* offer a valuable source of information for culinary historians with regard to such questions as appetite, digestion, the recommended quality and quantity of food, the frequency of meals, and the order of dishes, thus complementing the surviving cookbooks, which often contain very little information on these aspects of medieval culinary culture. Although it is clear that there initially was a strong connection between medical texts, books of manners and cookery books, Mennell (1986) argues that “cookery books rapidly emerged as a genre distinct from books about either medicine or manners” Mennell (1986: 66).<sup>8</sup> It seems that the practice of writing down recipes and collecting them into what could be called cookery books properly caught on in the 14<sup>th</sup> century, and by the 15<sup>th</sup> century the copying of recipe collections could no longer be considered unusual, suggesting an increasing interest in food in courtly circles (Scully 1992: 4; Strong 2002: 79). Although most of our surviving recipe collections come from this period, it seems likely that the roots of the knowledge codified in them goes back both to earlier written collections—now lost—and to an oral tradition which was passed on from mother to daughter and from master to apprentice (Redon, Sabban and Serventi 1998: 4).

Although some scholars of the past few decades (e.g. Mennell 1986: 49) have still perpetuated the myth about the scarcity of surviving medieval culinary collections, it is not true by any stretch of the imagination. Scully (1992: 24) has identified at least seventy-five medieval culinary recipe collections of European origin, representing the languages of most European nations and varying in size from a handful to three hundred recipes, the average being around a hundred. The oldest of these is held to be the work of German origin that Rudolf Grewe has named the *Northern-European Cookbook*, surviving in four distinct versions—two in Danish, one in Icelandic and one in Low German (Hieatt and Grewe 2001: 4–11)—and being “undoubtedly composed in the first half of the thirteenth century” (Scully 1988: 26). While a full survey of surviving continental recipe collections is beyond the scope of this thesis, the close relationship and common history of English and French cookery mean that the French tradition of culinary writing is an essential key to understanding and interpreting Middle English recipes (Hieatt

<sup>7</sup> According to Weiss-Amer (1992: 74), this was due almost exclusively to the influence of the so called ‘French School’ of dietetics, whose most influential works were the German *Sanitatis conservator* of Konrad von Eichstätt, compiled around 1300, and the *Regimen sanitatis ad inclitum regem Aragonum* of Arnald de Villanova, probably prepared at Montpellier in 1308 for king James II of Aragon (Weiss Adamson 1995: 110, 142). Konrad’s work—which served as a source for many later *regimina*—adjusted the culinary recipes of the Arabic masters to the palates and cooking practices of western Europe, leaving out most of the Arabic dishes and foodstuffs unavailable or unknown in France and Germany and bringing in new ingredients, such as almond milk (Weiss-Amer 1992: 77). Arnald de Villanova’s *regimen* included 18 recipes, and established “a close link between cooking and nutrition hitherto unknown to regimen-literature” (Weiss Adamson 1995: 116). The slightly later *Regimen sanitatis* of Magnus Mediolanensis already included 35 recipes—covering 48 pages—in addition to which both he and Arnald included an entire chapter on sauces (Weiss-Amer 1992: 74–75). The *Tractatus de regimine sanitatis* of Arnold von Bamberg went even further and includes a total of forty recipes, coming close to being “a cookbook organized according to the traditional dietetic division of foodstuffs into groups” (Weiss-Amer 1992: 77).

<sup>8</sup> Since the relationship between medical and culinary writing has not been investigated to a sufficient degree, the validity of this statement remains uncertain, and poses an interesting question for further study.

1998a: 137), and thus merits at least a brief mention of the three most significant French recipe collections: *Le Viandier de Taillevent*, *Le Ménagier de Paris* and *Du fait cuisine*.

Although *Le Viandier de Taillevent* has been attributed to the well known master cook ‘Taillevent’ (Guillaume de Tirel, cook to Charles V and Charles VI of France), it is now known that the earliest extant version of the collection was in fact written around 1300, at least 10 years before his birth. Thus what later went under his name, was in fact a later reworking of an existing collection, whose later incarnations differed quite significantly from the earliest one (Strong 2002: 78–79). *Le Ménagier de Paris*—which derives much of its material from the *Viandier*—was written by an author who was clearly educated, but most likely not of the royal court, since he explicitly refers to some dishes being suited only to the royal court and therefore beyond his means (Power 1992: 25). It has traditionally been considered to be written by a mature and wealthy Parisian bourgeois for the instruction of his young wife, but later research has argued its author to be a certain Gui de Montigny, who was in the service of the Duc de Berry (Strong 2002: 78–79). The *Ménagier* provides an interesting insight into the procedures of medieval cooking, since it describes aspects of cooking which are not usually described—probably because they were considered to be obvious by professional cooks (Scully 1988: 26–27). *Du fait cuisine*, on the other hand, was written in 1420 by Maistre Chiquart Amiczo, the *maistre queux* of Amadeus VIII, first Duke of Savoy and eventual Pope Felix V. According to Scully (1988), the special significance of this work lies in the fact that it “affords the most detailed view we presently possess of the techniques of food preparation in a late-medieval princely household” (27).

### 8.1.1 Surviving English collections

The first surviving work from Britain dealing with culinary matters is *De nominibus utensilium*, a 12<sup>th</sup>-century work written by an Augustinian canon, Alexander Nequam. It is not strictly a cookbook, but a glossary of everyday Latin words with glosses in English and Anglo-Norman containing some recipe suggestions (Goody 1982: 136).<sup>9</sup> Although Edwards and Pearsall (1989) are of the opinion that pre-1400 manuscripts containing vernacular secular works are merely “oddities and exceptions” (257), English recipes seem to have been written in the vernacular from the very beginning. Not surprisingly, considering the linguistic situation, the first proper recipe collections from England were written in Anglo-Norman, reflecting a culture with strong ties to France on the upper levels of society. The earliest of these is a collection of 29 recipes, found in BL MS Additional 32085 under the rubric *Coment l'en deit fere viande e claree*, which dates from the end of the 13<sup>th</sup> century (Hieatt and Jones 1986: 859; Scully 1988: 25), and the second is an unnamed collection of 32 recipes, found in BL MS Royal 12.C.xii, dated to between 1320 and 1340 (Hieatt and Jones 1986: 859–60; Scully 1988: 28).

Although the relationship of the later Middle English collections to early An-

<sup>9</sup> What makes these pseudo-recipes interesting is the fact that all of them—according to Hieatt and Butler (1985: 1)—appear in fuller form in later recipe collections. A linguistically interesting detail is the fact that oft-quoted tendency of the English language of using mostly Anglo-Saxon terms for live animals and mostly French terms for their meats is already visible here (e.g. cow/beef, calf/veal, deer/venison, sheep/mutton and pig/pork) (Goody 1982: 136).

glo-Norman and Continental French collections is unclear in most cases, we are fortunate in that the earliest surviving Middle English recipe collection, surviving in two manuscripts, BL MS Additional 46919 and BL MS Cotton Julius D viii,<sup>10</sup> and edited by Hieatt and Butler (1985: 43–58) under the title *Diuersa Cibaria* (DC), is in fact a 14<sup>th</sup>-century translation of these two Anglo-Norman collections (although the BL MS Add. 32085 collection is translated only in part).<sup>11</sup> In addition to being a translation of an earlier Anglo-Norman collection, the DC also contains recipes that have parallels in the earliest manuscript version of *Le Viandier de Taillevent* (Sion, Archives cantonales du Valais, S. 108), demonstrating a direct link to the continental French culinary tradition. As Meredith (2004: 29) has pointed out, the survival of such related collections in different languages is extremely valuable for comparative purposes as it affords us a glimpse not only of medieval translation practices in general, but also of the ways in which the largely French-based culinary lexis of English entered the language.

Unfortunately, while the later recipe collections of the 14<sup>th</sup> and 15<sup>th</sup> centuries also contain much that was borrowed from France, no direct links between the major 15<sup>th</sup>-century English and French recipe collections have been discovered (Scully 1988: 28, 1992: 220).<sup>12</sup> While clearly influenced not only by French but also Italian sources, both the early Anglo-Norman and later Middle English collections feature dishes which do not appear in any French sources, which has led Hieatt and Butler (1985: 6) to conclude that Anglo-Norman cooks also made significant original contributions to English cookery, which may even have been imitated on the continent.<sup>13</sup>

As was pointed out above, the first surviving recipe collections written in Middle English come to us from the 14<sup>th</sup> century. Contrary to the assumption of many scholars but in accordance with the general European situation, the number of surviving manuscripts containing culinary recipes in English is quite substantial: Hieatt (1992: 15) professes personal knowledge of over fifty such manuscripts and suspects that even more exist. Furthermore, she has identified seven ‘families’ of English recipe collections which survive in several more or less closely related copies, which means that they are likely to have circulated as more or less established collections:

- 1) *Diuersa Cibaria* (DC): London, British Library MSS Additional 32085 and 46919, Cotton Julius D viii, Royal 12.C.xii.
- 2) *Diuersa Servisa* (DS): Oxford, Bodleian Library MS Douce 257 (only complete exemplar).

<sup>10</sup> Individual recipes from these collections also appear in other manuscripts, including Oxford, Bodleian MS Laud Misc. 553 and Rawlinson D 1222.

<sup>11</sup> In addition to the collection being a translation from a known source, it is also known that the BL MS Additional 46919 was “compiled under the direction of, and partly in the hand of, Friar William Herebert of Hereford” (Hieatt and Butler 1985: 7).

<sup>12</sup> The influence of Anglo-Norman is also visible for example in the fact that the later Middle English collections preserve many of the Anglo-Norman titles for the recipes, which Wilson (2007: 44) sees as an indication that these collections were also originally translated from Anglo-Norman.

<sup>13</sup> In order to help trace the complex patterns of influence and interaction between different recipe collections, Hieatt (1992: 18) advocates the compilation for each edited collection of a list showing the order of recipes within it, allowing one to distinguish patterns of borrowing more quickly than comparing individual recipes. Such a list—or rather a series of lists—has also been drawn up for the *Potage Dyvers* collection and presented in section C.1.



- 3) *Utilis Coquinariorum* (UC): London, British Library MSS Sloane 468 and 374.
- 4) *Forme of Cury* (FC): Durham, University Library MS Cosin v.iii.11; London, BL MSS Additional 5016, Cotton Julius D viii, Harley 1605, Arundel 334<sup>14</sup>; Manchester, John Rylands Library MS English 7; New York, Pierpont Morgan Library MS Bühler 36; New York, Public Library MS Whitney 1<sup>14</sup>; Aberystwyth, National Library of Wales MS Peniarth 394 D<sup>14</sup>.
- 5) *Potage Dyvers* (PD): London, British Library MSS Harley 279 and 4016, and Additional 5467<sup>15</sup>; Oxford, Bodleian Library MSS Ashmole 1439 and Douce 55; Durham, University Library MS Cosin v.iii.11<sup>16</sup>.
- 6) *An Ordinance of Pottage*: London, British Library MSS Sloane 7 and 442; Oxford, Bodleian Library MS Rawlinson D 1222<sup>17</sup>; New Haven, Yale University MS Beinecke 163
- 7) *A Noble Boke off Cookry*: MSS London, Society of Antiquaries 287; Norfolk, Holkham 674; (Wiltshire, Longleat House, Pynson's *Boke of Cokery* (1599)) (Hieatt 1992: 21)

In identifying these 'family' groupings, the criteria used for identifying collections as parallel versions of each other are the similar, or identical, wording of the recipes, and "the way in which they occur in the same order – or, sometimes, roughly the same order – as groups" (Hieatt 2004: 27). These criteria are not always clear-cut, especially on the level of entire collections, as is demonstrated by Hieatt's (1996) ambivalence about whether to regard the different versions of the *PD* as separate collections or versions of the same collection: although she comments that the collection found in MS Additional 5467 has never been edited, she also notes that it "appears to contain substantially the same collection" (54) as the manuscripts edited by Austin (1888).

Perhaps the best known of these collections is the one known as *Forme of Cury*, which was contemporaneous with the French *Menagier* and was "compiled of the chef Maister Cokes of kyng Richard the Se[*cu*]nde kyng of [En]glond aftir the Conquest . the which was accounted þe best and ryallest vyaund[*ier*] of alle cristen [k]ynges" (Hieatt and Butler 1985: 20; Strong 2002: 79).<sup>18</sup> Although it was not the largest collection of recipes, its longest versions containing about 200 recipes, it seems to have been central to the tradition, since almost all of the larger 15<sup>th</sup> century English collections contain groups of recipes belonging to it (Hieatt 1992: 16). Hieatt (1992: 16) confers the honour of the largest Middle English recipe collection on the collection found in Holkham MS 674 and printed by Napier (1882) under the title *A Noble Boke off Cookry*, although at least Napier's edition only contains 251 recipes, which is 18 less than MS Ashmole 1439, one of the six *PD* MSS edited here.<sup>19</sup> In terms of known surviving manuscript versions, the most widespread of these families seems to have been the *FC* with its nine known versions. When editing the *Ordinance of Pottage* collection in 1988, Hieatt considered it to be "one

<sup>14</sup> This MS version is not listed in Hieatt (1992) but has been added based on Hieatt and Butler (1985)

<sup>15</sup> This MS version is not listed in Hieatt (1992) but has been added based on Hieatt (2004).

<sup>16</sup> This MS version is not listed in Hieatt (1992) but has been added based on Marttila (Forthcoming).

<sup>17</sup> This MS version is not listed in Hieatt (1992) but has been added based on Hieatt (1988).

<sup>18</sup> MS Cosin V.III.11, one of the manuscripts containing a version of the *PD*, also contains a copy of this recipe collection (see Hieatt and Butler 1985: 20–30, and appendix F of this thesis).

<sup>19</sup> The number of recipes contained in the different Middle English manuscripts containing recipes varies enormously from a few odd recipes added to an empty margin to the full-blown collections listed above, containing 20–250 recipes (Hieatt 1992: 16).

of the longest, most widely disseminated and most typical of 15<sup>th</sup>-century English culinary collections" (9), being second only to the *Forme of Cury* with its four surviving manuscript versions (with some of the recipes also found in other collections). However, as has been noted in Hieatt (2004) and Marttila (Forthcoming) and will be argued in chapter 9, the *PD* family of collections with its six surviving manuscript versions could be seen as an even more widespread one, although it could be argued that these six manuscripts in fact constitute anything from two to five different collections (see chapter 13).

In addition to these 'established' families of collections, there also exist a number of manuscripts—many of which have been described and edited for their previously unedited parts in Hieatt (2008)—that contain material from several of these collections and from other, hitherto unidentified sources. An example of such a collection, combining material from several of the families defined by Hieatt (1992) with recipes that are not attested elsewhere, is found in British Library MS Harley 5401 and edited in Hieatt (1996). This relatively short collection of 96 recipes combines groups of recipes from the *FC*, the *DS*, and the *PD* families, along with individual recipes from also other sources, and often provides clearer and less corrupted readings of them than the manuscripts belonging to these collections. In fact, many of the larger surviving English recipe collections—e.g. National University of Wales MS Peniarth 394D, Cambridge University Library MS L1.I.18 and B.L. Arundel 334—contain material from several of the aforementioned families, and very few manuscript collections of significant size are entirely independent of them (Hieatt 1992: 17-8).<sup>20</sup>

### 8.1.2 Existing editions and other research tools

In addition to having provoked a vast amount of secondary literature, mostly written from the point of view of culinary and cultural history, Middle English culinary texts have also been edited quite extensively, although a many of the established editions are from the 19<sup>th</sup> century and therefore do not necessarily fulfil the standards of modern editorial practice (see e.g. Morris 1862, Napier 1882, and Austin 1888). Of modern editors and scholars of culinary recipes, Constance C. Hieatt (see e.g. 1992, 1998 and Hieatt 1998b) has been the most active proponent of new editions of Middle English culinary recipes and written extensively on the various aspects of editing them. An invaluable tool for an editor is constituted by the "Répertoire des Manuscrits Médiévaux Contenant des Recettes Culinaires" (Hieatt, Lambert et al. 1992), which provides a listing of all European manuscripts known to contain culinary recipes, and included information on any editions based on

<sup>20</sup> However, an interesting example of such an independent collection, found in Cambridge Corpus Christi College MS F 291, has been identified and edited by Hieatt (2012) and does not have any identifiable sources. According to Hieatt (2012: 22) she and the late Sharon Butler discovered and started to transcribe the collection already in 1980, but because of its considerable textual problems, they decided against editing it at the time and as a result the edition was only published in 2012. This collection is interesting in the sense that the recipes in it seem to exhibit the features of late 15<sup>th</sup>-century recipes "to an extreme degree" (Hieatt 2012: 14): "no other elaborates its recipes to such great length, or gives do many exact measurements of ingredients, or goes further in sweetening its meat and fish recipes with dried fruits and other sweeteners, although some go almost as far in the latter respect" (14-5). For this reason, Hieatt characterizes it as "the ultimate fifteenth-century cookery book" which "carries all the tendencies of recipe writing of its period about as far as they can go" (15).

them up to 1992. The “Répertoire” lists the manuscripts organized by the city and library in which they are held, and provides not only the library shelfmark and folio numbers containing the recipes, but also information on the language and possible title and author of the culinary section, the number of surviving recipes, a dating and a localization of the manuscript when possible, and a list of all modern editions, translations and catalogue descriptions pertaining to the recipe collection. For English recipes, they list a total of 40 manuscripts, including the two manuscripts containing recipes in Anglo-Norman mentioned above, with some manuscripts (like the DUL MS Cosin V.iii.11) containing several independent collections. While all of the manuscript edited in the present edition are included in the list, not all of the recipe collections (i.e. the MS C and MS Ad) are correctly identified as members of the *Potage Dyvers* family.

Organized by works rather than individual manuscripts, also the *Index of Middle English Prose* (Keiser 1998a) lists 12 separate culinary recipe collections, several of which have been edited more than once (3887-91). The earliest of the modern editions listed by Keiser (1998a: 3887) is Samuel Pegge’s 1780 edition of the *FC*, based on BL Additional MS 5016 (his edition also included a version of the *DS*, based on Bodleian MS Douce 257).<sup>21</sup> This collection—perhaps the most famous of the English collections—was edited from the British Library MS Arundel 334 version by Nichols in 1790 as a part of his *Collection of Ordinances and Regulations for the Government of the Royal Household*. In 1791, Warner published another edition of the MS Additional 5016 version, which was essentially a republication of Pegge’s edition. In the 20<sup>th</sup> century, the *FC* was first published as a critical edition in a dissertation (Sass 1979), before the latest and ‘definitive’ critical edition by Hieatt and Butler (1985: 93–145).

Other 19<sup>th</sup>-century editions of Middle English culinary manuscripts include Morris’s (1862) edition of the culinary text in verse known as the *Liber Cure Cocorum*, based on British Library MS Sloane 1986, Napier’s (1882) edition of the *Noble Boke Off Cookry*, based on “a rare MS. in the Holkham Collection”,<sup>22</sup> and Austin’s (1888) edition of BL MSS Harley 279 and Harley 4016—which are also edited in the present edition—published under the title of *Two fifteenth century cookery books*. While Austin’s edition is not suitable for linguistic use since it supplies editorial punctuation, expands abbreviations, and frequently divides or combines compound words according to modern practice, it is not a critical one and thus presents reasonably accurate transcriptions of the two manuscripts, with some variants from the other two presented in footnotes and in a separate collation list for MS Ashmole 1439 (Austin 1888: xviii-xix). Unfortunately it does not preserve the original lineation (although page breaks are indicated by footnotes) or any other features of the layout or decoration of the manuscripts. Its most serious shortcoming, however, is its total lack of editorial documentation, typical to the editions of its time; its preface contains no mention of its editorial principles and practices, focusing instead on the historical background of the various feasts described by the included menus. All that is mentioned of the original sources and their treatment is that the two main manuscripts (British Library MSS Harley

<sup>21</sup> Incidentally, this edition is also the earliest *type facsimile* edition produced of a Middle English text (Edwards 1987: 45, 2000: 72)

<sup>22</sup> This text also survives in a printed version from around 1500, printed in London by Richard Pynson with the title *This is the boke of cokery*.

279 and 4016) were “collated with” the secondary ones (Bodleian Library MSS Ashmole 1439 and Douce 55), leaving the activities of the editor to be inferred through a comparison of the transcriptions and the original sources, not possible for most users of the edition.

Despite the increasing popularity of the topic in recent decades, the number of new editions in the late 20<sup>th</sup> century is surprisingly low, and even the ones published are very traditional in their editorial approach. Apart from reading edition reprints of out-of-copyright 19<sup>th</sup>-century editions<sup>23</sup>, the first edition of new material was G. A. J. Hodgett’s 1972 facsimile edition—with a modern English translation—of the recipe collection known as *Stere Htt Wele*, found in Cambridge Magdalene College MS Pepys 1047. While such a facsimile edition is of limited use to linguists and other scholars, it does provide a rare glimpse at the layout and other visual features of a late medieval utilitarian manuscript. In 1976, Hieatt and Butler published a selection of modern adaptations of medieval recipes from earlier published editions, including Pegge (1780), Morris (1862), Austin (1888), Napier (1882), and Power (1992, originally published in 1928),<sup>24</sup> and in 1979 Sass prepared a critical edition of the *FC* for her PhD thesis, after also having published a book of modern adaptations based on its recipes in 1975.

In 1985, Hieatt and Butler prepared a new, critical edition of not only the *FC*, but also of three other 14<sup>th</sup>-century recipe collections—titled *Diuersa Cibaria*, *Diuersa Servisa* and *Utilis Coquinario* in the list above—and a selection of miscellaneous recipes. It was published in the Early English Text Society series and became not only the ‘definitive’ edition for these collections but also the best-known and most widely-used scholarly edition of medieval culinary recipes, used by a large number of scholars—including linguists—up to the present day (see e.g. Görlach 1992, 2004, and Meredith 2004).<sup>25</sup> In addition to the *FC*, the 1980s also saw an edition of the two earliest known English collections mentioned above, namely the Anglo-Norman culinary recipe collections found in BL MSS Additional 32085 and Royal 12.C.xii by Hieatt and Jones (1986), and a critical edition of the family of collections known as *An Ordinance of Pottage*, based on all the four known manuscripts with the Yale University Beinecke MS 163 as the base text by Hieatt (1988).

During the 1990s, the only new edition of culinary recipes seems to have been Hieatt’s 1996 single manuscript edition of the unnamed short recipe collection found in MS Harley 5401, which provides a reading version of the text of the recipes, with editorial punctuation, normalization of capitalization and silent expansion of abbreviations.<sup>26</sup> In addition to this new edition, the 1888 edition of

<sup>23</sup> Such as *A Fifteenth Century Cookry Boke* (1962), which merely presents the recipes originally edited by Austin in a new order in a clear-text format with a simplified and more explanatory version of Austin’s Glossary.

<sup>24</sup> A 2<sup>nd</sup> edition, revised by Hieatt and Brenda Hosington, was published in 1996.

<sup>25</sup> The popularity of the edition as material for linguistic study is somewhat surprising, since as Carroll (2009: 60) points out, it is a critical edition with all of the attendant problems outlined in section 3.1. Already in his 1988 review Keiser noted that some of the decisions made by the editors “will not be satisfactory for all” (411), historical linguists in particular. Especially the decision to incorporate into the base manuscript entire recipes from other manuscript versions is problematic in terms of the integrity of its textual structure and the variation inherent in its different versions.

<sup>26</sup> The edition also provides notes on the parallel versions of each recipe found in other published editions; these have been used, together with the *Concordance of English Recipes* (Hieatt and Nutter

Austin was given new currency, especially among amateur culinary historians, by Renfrow's 1990 translation and modern adaptation of the recipes contained in it (second edition published in 1997).<sup>27</sup> Editions of Middle English culinary recipes published over the last decade include Hieatt (2008), which edits "the contents of three fairly brief collections which have never been edited before, recipes from several others which were previously used for collation (or could have been) but contain other recipes not previously noted, and a few recipes which occur in isolation or in very small groups" (Hieatt 2008: 9), and Hieatt (2012) which contains an edition—with translations and commentary—of the unique copy of a late 15<sup>th</sup>-century recipe collection containing very detailed recipes, found in Oxford Corpus Christi College MS F 291. The latest ME recipe edition to be published is Marttila (Forthcoming), which will present a contextualized diplomatic edition of the version of *PD* contained in MS C, which is also included in the present digital edition.

Unfortunately for historical corpus linguists, all of the existing editions of Middle English culinary recipes are traditional printed editions.<sup>28</sup> While most of the 18<sup>th</sup>- and 19<sup>th</sup>-century ones—and those newer ones editing a unique manuscript—reproduce the text of a single manuscript copy and are thus in principle suitable for linguistic research, the documentation of their editorial practices is severely lacking and it is thus impossible to know the kinds of emendations or alterations undertaken by the editor.<sup>29</sup> Furthermore, even the more diplomatic of the editions do not reproduce the layout or other paratextual features of the manuscripts, instead conforming them to the traditional format of the printed edition. Being geared mainly towards the needs of culinary historians, most of the 20<sup>th</sup>-century editions are critical ones that combine readings from several manuscripts or at least emend the text, violating not only the paratextual but also the textual integrity of the original documentary text and resulting in many of the problems outlined in section 3.1.

In addition to the editions and catalogues of culinary manuscripts, an extremely valuable resource for textual scholars of historical recipes was produced by Hieatt and Nutter in the form of the *Concordance of English Recipes: Thirteenth through Fifteenth Centuries* (2006) (from here on referred to as the *Concordance*). It lists all of the recipes from all editions published up to 2006, organized according to their lemmatized name—a generalized modern English version of the name or a descriptive title—collecting together not only the different parallel versions of one recipe but also recipes describing different versions of the same general dish. This concordance is used for two purposes in this thesis: first of all, in the edition itself

---

2006) and its supplements (in Hieatt 2008 and 2012), to associate the recipes of the *PD* family to their parallel versions in other collections in section 9.3.

<sup>27</sup> From the point of view of historical or linguistic scholarship, this republication does not add much value, considering that in addition to Austin's editorial misinterpretations, 's inability to let go of present-day American culinary tastes adds a host of new misinterpretations to the translations and adaptations.

<sup>28</sup> As an exception, although not precisely an edition, Manchester University Library does provide transcriptions of the contents of all the folia the MS English 7 version of *FC* along with the high-quality facsimile images contained in the digital Rylands Medieval Collection (<[http://enriqueta.man.ac.uk:8180/luna/servlet/Man4MedievalVC\\_4\\_4](http://enriqueta.man.ac.uk:8180/luna/servlet/Man4MedievalVC_4_4)>).

<sup>29</sup> For example Austin (1888), while seemingly offering a diplomatic transcription, complete with indications of abbreviations, flourishes and other palaeographic detail, simultaneously emends the base texts (Harley 279 and Harley 4016) with readings from Douce 55 and Ashmole 1439, and also contains a significant amount of transcription errors.

to provide a lemmatized name (or several in some cases) for each *recipe type*<sup>30</sup> of the *PD*, and second, in chapter 9 to associate each of these recipe types to their parallel versions found in the collections covered by the *Concordance*.<sup>31</sup>

## 8.2 Medieval recipe collections as textual objects

From a textual point of view, recipe collections—and other structurally similar texts like collections of statutes or encyclopædias—can be viewed on two different levels: either as a single large entity made up of short, independent segments—a *discourse colony* as described in subsection 2.1.2—or as a group of single independent entities that simply happen to share a physical location, each of these views representing a different type of textual object. While the appropriate viewpoint ultimately depends on the kinds of questions we want to ask, most studies of medieval recipes and other instructional material have focused predominantly on the individual recipe, mainly because the recipe collection as a discourse colony does not readily agree with the conventional concept of *text* used by text linguists and textual scholars (Carroll 2003: 149-50, 2006: 314), and there is often considerable uncertainty whether two collections containing largely the same recipes should be considered distinct works or merely different manuscript versions of the same work. In comparison, an individual recipe has commonly been seen as a much more convenient unit of analysis, allowing for the easy comparison of one token to another.

As was already pointed out in subsection 3.1.1, this means that the textual identity of a recipe collection is a second-order phenomenon: it needs to be established first on the level of individual recipes, i.e. whether the individual recipes contained in two collections represent independent *works* or are sufficiently similar to be considered parallel textual realizations of the same work, and then on the level of the collection, i.e. whether two collections contain a sufficient number of parallel recipes and whether their order is sufficiently similar for them to be seen as two textual realizations of the same ‘compound work’ or discourse colony, or do they represent individual works drawing on common material. The first of these levels is considerably simpler. In the present edition, the definition of a *parallel recipe* is similar to that of Hieatt (2004), summarized in subsection 8.1.1 above, but in accordance with the distinction between *work* and *version* outlined in section 2.1, emphasizes the semantic content of the recipe instead of its linguistic surface realization: parallel recipes are here defined as recipes that describe the same procedure using mainly the same ingredients, with only minor variation in detail. As to the second level, the complexity of the relationships between entire collections

<sup>30</sup> The term *recipe type* is here used to collectively refer to all of the parallel versions of what is considered to be ‘the same recipe’ (see section 9.1).

<sup>31</sup> It is unfortunate that the *Concordance* has been published as a printed volume instead of an online resource, since a database containing all of the data in the concordance in digital form would perform the second of these operations automatically based on the first and allow the formal and persistent linking of parallel versions of recipes in different manuscripts. The creation of an electronic database containing the data of the *Concordance* would be trivial from a technical standpoint (and based on the introduction, the published version was in fact based on such a database), but unfortunately copyright issues—complicated by the fact that the copyright of the *Concordance* is held by the Arizona Board of Regents for Arizona State University—are likely to make the Open Access online publication of such a database difficult.

means that the absolute concept of a 'parallel version' is hardly appropriate. While the initial postulate of this thesis is that the six recipe collections edited here are more closely *related* to each other than to other surviving recipe collections, the analysis undertaken in chapter 13 also shows that these relationships vary considerably between different pairs of versions and with respect to different parts of the collection.

As an edition of six quite different manuscript versions of what can on the level of a work be considered the same discourse colony, this thesis aims to enable the study of medieval culinary recipes both on the level of individual manuscript copies as independent discourse colonies, and on the level of individual recipes within the collection. A traditional critical edition of several manuscript copies—which would in fact represent an entirely new *discourse colony*, separate from any of the medieval ones—would compress together the features—both textual and codicological—of its source collections, thus precluding any comparative study of their differences and similarities on the level of the collection as a whole. As all students of medieval food quickly discover, the vast majority of medieval recipes are repeated in collection after collection (Hieatt 1992: 18). From examining different recipe collections—and different versions of the same collection—it seems clear that recipes travelled extensively in both time and space,<sup>32</sup> as successive recipe collections were enriched by borrowings and adaptations as new generations of cooks first absorbed the traditional rudiments of the craft during their apprenticeship and then grafted onto this tradition their own innovations (Scully 1992: 197–198).<sup>33</sup>

Thus, contrary to the precepts of traditional stemmatic bibliography, the numerous similarities apparent in cookery manuscripts are not likely to be entirely due to some common ancestor collection, but also due to the fact that the cook learned his trade through apprenticeship and the craft was primarily transmitted orally (Scully 1992: 196). This means that the links between different versions of recipes are not only intertextual but also interpractical: while two versions of a recipe may well have been written down completely independently of each other, the underlying practices—the usage of ingredients and methods—are essentially the same, resulting in parallel recipes of similar content but varying linguistic realization (Scully 1992: 197). Thus the traditional tree metaphor of textual relationships that has dominated 19<sup>th</sup>- and 20<sup>th</sup>-century textual editing seems to bear little resemblance to the reality of medieval culinary recipes, and its replacement by the rhizome (or grass) model is especially welcome in their case (Greetham 1998: 299).<sup>34</sup>

<sup>32</sup> Although the exact same collection of recipes does occasionally occur in several manuscripts, it is more common to find 'families' of collections, which contain much of the same recipe material but differ in its organization or presentation (Redon, Sabban and Serventi 1998: 2).

<sup>33</sup> It was this inherently conservative and evolutionary rather than revolutionary nature of medieval cookery that gave the recipes a very long life span (Wheaton 1983: xxi) and to a great extent served to promote the development of a unified culinary culture across Europe.

<sup>34</sup> Although in the case of culinary recipes and other practical texts, the model should be furthermore extended to cover also the oral transmission of texts between manuscript copies.

### 8.2.1 Recipe collections as *discourse colonies*

Following Carroll (2003), this thesis takes the view that recipe collections are best seen as *discourse colonies*, or collections of “essentially self-contained units which have been compiled into a larger entity” which can in some ways be treated like a single text (e.g. provided with a title and catalogued as an entity), but differ from a narrative or prototypical texts in the sense that they are not intended to be read continuously from beginning to end and can have their component parts ordered differently in different manuscript versions (Carroll 2003: 138 and 2006: 314). In terms of the nine characteristics of discourse colonies defined by (Hoey 2001: 88) and outlined in subsection 2.1.2, Carroll (2006: 315) has found medieval culinary recipes to possess either eight or all nine of them, which makes them even more archetypal as discourse colonies than modern recipe collections, which Hoey (2001) found to exhibit only six or seven of the characteristics. While Hoey’s (1986 and 2001) analysis of present-day discourse colonies does not in any way take into account their visual paratext, such as typography or page layout, such features are—both in medieval and modern texts—frequently used to highlight certain items (like titles or headings) and thus facilitate finding individual items within the colony (Carroll 2006: 320). In her analysis of the visual features of manuscript recipe collections, Carroll (2006) has found also visual manuscript evidence to support the analysis of medieval recipe collections as discourse colonies, many of the characteristic features of discourse colonies being signalled by visual features of the manuscript.<sup>35</sup>

While medieval culinary recipes fulfil the first characteristic of discourse colonies in that the meaning of neither the individual recipe nor the collection as a whole is affected by the sequence of the recipes and individual recipes function as self-contained and freely mobile units—as is demonstrated by the considerable variation in the ordering of the different *PD* versions (see chapter 13)—this does not mean that they are always completely isolated. As Carroll (2003: 153, 2006: 318) points out, some deictic cross-references to other recipes in recipe collections do occur, but since they are not inherently stronger between adjacent members of the colony than they are between non-adjacent ones, they do not constrain the reading order (Hoey 2001: 74). The insignificance of the order of the recipes within the collection also means that adjacent recipes are not connected to each other by any kinds of cohesive devices and do not form continuous prose, also fulfilling the second characteristic. The isolated nature of individual recipes is also reflected on the level of discourse structure, with each recipe in a collection repeating the same highly formulaic internal textual structure consisting of a clearly marked beginning and end.

Hoey’s third feature, the framing context, is slightly problematic in the medieval context, as works are not always given titles, even when they are understood as entities. However, as Carroll (2006: 317) points out, the framing context in medieval recipe collections may also be provided by an incipit (effectively serving as a title), a preface, or a table of contents. For example in the case of *PD*, none of the versions give a proper title for the entire collection, but all versions except

<sup>35</sup> Carroll (2003: 152) sees for example the presence of rubricated initials at the beginning of each recipe and the fact that each recipe begins on a new line as emphasising the unitary and unconnected status of individual recipes in a collection.



for MS C have some kind of an incipit or a heading characterising the collection (or its component parts) as having to do with food—the most prominent being the running heads at the top of every page of MS H279.<sup>36</sup> However, the strongest framing context in the case of the *PD* is provided by the table of contents included in all versions except for MS H4016, explicitly enumerating the recipes making up the collection.<sup>37</sup> However, even in the absence of these, the prototypical layout and visual organization of a recipe collection may be considered to provide a sufficient framing context, as it enables the reader to immediately recognize the text type. Hoey's fourth characteristic, attribution to several authors or anonymity, is in fact not usually shared by modern cookery books (Hoey 2001: 82), but is very typical of Middle English recipe collections, none of which are attributed to a named author.<sup>38</sup>

In terms of the fifth characteristic of discourse colonies, i.e. the possibility of non-sequential and selective reading, culinary recipes are very prototypical, being intended precisely for this kind of reading. Within the collection, individual recipes are self-contained not only in terms of their textual and linguistic structure, but also in terms of their semantic content, each recipe providing instructions for a separate process that is independent of the instructions and processes described in other recipes. The only exception to the independence of individual recipes is constituted by the occasional deictic cross-references between recipes, where one recipe refers to instructions provided in another, they can be considered to be the exception and thus not to undermine the status of the collections as a discourse colony Carroll (2003: 153). Examples of these kinds of inter-recipe deictic references in the *PD* versions include the recipes for a meat and fish gelée, of which the latter refers to the former by instructing the reader to do “as you do for the gelée of meat”.<sup>39</sup> Similarly, recipe PD 82 (Mortrews of Fish), which appears in all six versions, instructs the reader to serve it in the manner of the more common meat version (or more generally ‘other mortrews’) despite the fact that it is preceded by a recipe for the meat version only in MSS Ad and D.<sup>40</sup>

The applicability of Hoey's sixth and seventh characteristics, i.e. the reproducibility of individual members in another colony and the ability to add members

<sup>36</sup> However, some versions provide parts of it with amounts to titles in the modern sense (the title *Potage Dyvers*, which has become associated with the entire collection, is in fact the title of the first part of the MS H279 version, displayed prominently at the head of every page.

<sup>37</sup> In MS C, the table of contents is the principal indication that the incomplete ending of another recipe collection following the *PD* collection on ff. 22-25 is in fact a part of a separate collection whose beginning has been lost with some missing folia and not the end of the *PD* collection. On the other hand, the table of contents included in MS As has for some reason been left incomplete by the scribe, covering only about half of the entire collection, being followed by four empty folia.

<sup>38</sup> The reference to the ‘master cooks of Richard II’ in the *FC* is the closest an English recipe collection comes to such an attribution. This is an interesting difference to continental recipe collections, several of which are attributed to a named individual, as was pointed out above.

<sup>39</sup> In MSS Ad, H279, As and D these recipes are consecutive, the fish one following the meat one, but in MSS C and H4016 they are separated by over 30 other recipes.

<sup>40</sup> In its most extreme form, this linking can result in a recipe consisting exclusively of such a link, as in the case of recipes PD 258 (Primrose Pottage) and PD 259 (Hawthorn Pottage), whose only content is an instruction to prepare it exactly as the recipe PD 256 (Violet Pottage). The frequent instructions to truss or carve a fish or fowl in the same manner as another do not really fit to this category, since in the majority of cases, the recipe ostensibly referred to does not actually contain the instructions, which are instead understood to be a part of the professional knowledge of the cook.

to a colony or omit them from it, is clearly demonstrated by the existence of collections combining recipes from several different families, and by the fact that the different versions of the *Potage Dyvers*—and other collections—contain different selections of recipes.<sup>41</sup> In the context of medieval textuality, however, this feature is not limited to what Hoey considers discourse colonies, as medieval compilers and collectors of texts in general seem to have “felt little compunction when excerpting and rearranging material from other texts”, the criterion for borrowing material being “not the integrity of the text, but its utility in context” (Gillespie 1989: 326).<sup>42</sup> An interesting example of a recipe collection explicitly designed as extensible is provided by the MS Rawlinson 1222 version of *An Ordinance of Potage*, which is organized into sections containing different types of dishes, each of which is followed by two or three blank pages which Hieatt (1988: 17) interprets to be intended for adding new recipes.

The eighth characteristic of prototypical discourse colonies, i.e. that many or all members have a matching functional relation with each other, is obviously true of recipe collections, as the recipes making up a collection can be considered to be identical in their function of providing instructions for preparing a dish and thus to have a parallel rather than serial relationship to each other. As Carroll (2006: 319) points out, this identity of function is also reflected in the repetition of similar linguistic structure and visual layout from one recipe to the next, resulting in a distinctively repetitive textual structure. The last characteristic of recipe colonies, the alphabetic, numeric or temporal ordering of the components, is the only one that Carroll (2006: 319) considers not to apply to the *DS* collection she has analysed, as she sees its recipes to be presented in a seemingly arbitrary order that does not help the reader in locating an individual item. In a sense this criterion—or at least Carroll’s interpretation of it is somewhat strange, since for any textual components that are not naturally ordered, including recipes, a numeric order is essentially an arbitrary one. But if we accept any numbering as fulfilling the criterion, most versions of *Potage Dyvers* fulfill also this requirement, as MSS Ad, D and H279 all have original recipe numbering, which is furthermore linked to a similarly numbered table of contents and clearly intended as a reference aid, while MS C has recipe numbers added in pencil by an 18<sup>th</sup>-century librarian and MS As has a numbered table of contents which can be seen to indirectly associate numbers with the recipes.<sup>43</sup>

Based on the criteria suggested by Hoey (1986, 2001), the *PD* recipe collection—as well as many other collections—would seem to fulfill all nine of the criteria at least to some degree and to represent a prototypical example of a discourse colony. While the discourse colony can thus be considered an accurate model of the recipe collection as a textual object, it does not provide straightforward answers to the problem of relating the ‘micro-level’ of the individual recipe to the

<sup>41</sup> For example Carroll (2006) points out that many of the recipes from the *DS* are found, together with recipes from the *FC* collection, in MS Harley 5401 (edited and described by Hieatt (2004)), and New York MS Whitney 1 (ff. 12-14v) contains a subset of the recipes, forming a shorter version of “the same discourse colony but with a smaller ‘population’” (318-9).

<sup>42</sup> For example sermon manuscripts—another example of discourse colonies—often freely adapt borrowed sermons by adding and removing material without any indication of the change of author (Gillespie 1989: 326).

<sup>43</sup> These manuscripts are far from being unique in this respect, as Carroll (2006: 319) points out that at least the MS Bühler 36 and MS Additional 5016 versions of the *FC* contain original recipe numbers.

‘macro-level’ of the collection for evaluating the textual relationships between different versions of the collection. The variability of structure and extent allowed by the characteristics of discourse colonies does, however, allow us to consider the six versions of the *PD* as variant versions of the same work despite the fact that some of the versions—if considered as a traditionally structured continuous texts—bear little resemblance to each other.<sup>44</sup>

### Organizational principles of recipe collections

As was pointed out above, a numerical ordering of inherently non-numerical items like recipes is an arbitrary one and cannot properly be considered an organizational principle, the numbering being more accurately described as a finding aid that can just as well be added to the recipe collection after its order has been established, as in the case of MS C. Although the organization of medieval collections often seems to follow no discernible logic when compared to the systematic organization of modern cookbooks, the order of recipes is rarely completely arbitrary, and modern scholars should be careful not to judge medieval collections by modern standards and thus miss what organization there is (Carroll 1999: 28). Based mainly on Continental examples, Redon, Sabban and Serventi (1998: 2) have distinguished three different principles of organization found in medieval cookbooks. The first, *alphabetical*, is unusual but does occur for example in the unedited collection contained in MS 136 of the Medical Society of London (Carroll 2006) and the Italian *Libro di cucina del secolo XIV* edited by Frati in 1899. The second is according to the *principal ingredient* (vegetables, meat, fish, eggs, etc.) of the dish, forming “a kind of culinary encyclopaedia” Redon, Sabban and Serventi (1998: 2), exemplified by the Latin 14<sup>th</sup>-century culinary treatise known as *Liber de coquina* and edited by Mulon in 1971. The third is according to the *type of dish* (roasts, pottages, sauces, fritters, pies, etc.), as in the *Libro de arte coquinaria* of Maestro Martino, edited by Faccioli in 1966. It must, however, be noted that the last two of these are rarely applied with perfect consistency, since dishes can have both several principal ingredients and belong to several types or categories, requiring a hierarchy of categorizations.<sup>45</sup>

None of these three principles seem to be followed very consistently in English collections. Hieatt (1988: 16-7) has argued that many Middle English recipe collections, for example the *FC*, are in fact organized according to the order in which

<sup>44</sup> It does also provide us with a theoretical basis for evaluating the similarity of whole collections through quantitative methods by separately calculating the similarities on the word-level—using a suitable edit distance metric such as Damerau-Levenshtein (Damerau 1964; Levenshtein 1966) or Jaro-Winkler (Jaro 1989; Winkler 1990) distance—of individual pairs of parallel recipes (identified either manually or through the use of similarity metrics) and combining the aggregate value of these distances with a similar distance calculated on the level of whole recipes (considering parallel versions as matches). These kinds of quantitative similarity metrics could then be used to support and guide cladistic analyses on the level of both individual recipes to determine the most likely transmission history of an individual recipe based on its linguistic features, and of whole collections to determine the transmission history of the organizational structure of the discourse colony, which may well differ from that of its individual component recipes.

<sup>45</sup> For example the recipes of *Le Viandier de Taillevent* are organized according to a primary division into dishes of land animals and fowl on the one hand, and fish and other aquatic creatures on the other. Further subdivisions are then made “according to the circumstances of their use and according to the methods employed in their cooking and preparation” (Scully 1988: 17).

foods were customarily served at English feasts, namely beginning with simpler and more substantial everyday dishes such as different meat and vegetable pottages, progressing to more delicate dishes like roasted birds, and concluding with sweeter tarts, fritters, cooked fruit and spiced wines (Hieatt 1988: 16-7, 2012: 11).<sup>46</sup> According to Hieatt (1988), this ordering of recipes confirms “the existence of a perfectly rational serving order in medieval England”, namely “the hearty basics first, and then—for those lucky enough to be served further courses—more interesting dishes with the rarer delicacies and dainties saved for the end of the meal”. However, she also notes that there are some collections that form an exception to this and truly do not seem to follow any discernible logic in their ordering, like the Corpus Christi College MS F 291 collection edited in Hieatt (2012), which has some “clumps of more-or-less related recipes” but “no discernable overall rationale and no resemblance to the order of any other collection” and in fact begins with one of the traditional ‘subtleties’, known as ‘cockatrice’ and consisting of the combined halves of a chicken and a suckling pig (11).

Within the *PD* family, the organization of the recipes varies widely between the versions, effectively dividing the six versions into three pairs, each of which have a distinctive order quite different from the others, although there are sequences of recipes that occur as units across more than two collections (see chapter 13 for a comparative analysis of the textual structure of the six versions). None of the versions seems to follow the ordering suggested by Hieatt (1988, 2012) to any degree of consistency, although all of the versions begin with a selection of ten or so recipes for pottages or boiled foods that could be characterized as “hearty basics”, but the only version that ends with anything that could be considered as dainties—namely pies, pastries and crispels—is MS H279, the other versions ending with a sequence of recipes for fish (MSS C and H4016), for sauces (MS As), or for dishes with no apparent shared features.<sup>47</sup> The majority of the versions—MSS Ad, As, D and H279—are divided into several ‘subcollections’ defined on the basis of various criteria, including the method of preparation (e.g. roasted or oven-baked dishes), the category of main ingredient (e.g. fish dishes) and the time of serving (e.g. Lenten or fish-day dishes), exemplifying the second and third principles outlined by Redon, Sabban and Serventi (1998).<sup>48</sup> Although not explicitly divided into subsections, also the two remaining versions, MSS C and H4016, seem to exhibit similar groupings of recipes, such as recipes for meatless dishes, for various fishes and wildfowl, and for roasted dishes of different types.<sup>49</sup> As Carroll (2006) has observed, these kinds of divisions, especially the separation of fish recipes from meat ones in a section of their own serve a pragmatic purpose: since the medieval reader was normally looking for either a meat dish or a fish dish (depending on

<sup>46</sup> Lehmann (2003) has observed this to hold also for the *Ordinance of Pottage* family of recipe collections, where “everyday dishes” come first and are followed by “the more ‘curious’ dishes” (24).

<sup>47</sup> For example the recipe for *cockatrice*, mentioned above, which should occur towards the end according to the order outlined by Hieatt, occurs quite near the beginning (18 % into the collection) in MS C and just after the middle in the other versions it occurs in (56–70 % into the collection).

<sup>48</sup> This kind of division—including the types of categories—also occurs on the Continent, for example *Le Viandier de Taillevent* containing a remarkably detailed division into subsections explicitly titled as: *Bouillitures*, *Potages lians*, *Rostz de chair*, *Entremés*, *Potages lians sans char*, *Pour malades*, *Chapitre de poisson d’eau douce*, *Chapitre de poisson de mer ront*, *Chapitre de poisson de mer plat*, *Viande de Quaresme*, *Saulces non boullues* and *Saulces boullues* (Scully 1988: v).

<sup>49</sup> The organization of recipes within the six versions will be discussed in more detail in section 13.3.

the day), he or she could easily focus his or her attention to the relevant section of the collection.

### Recipe collections as reference works

Like many texts characterized as discourse colonies, recipe collections were most likely not intended to be read sequentially but to be used for pragmatic *reference reading*, where the reader is interested in a specific piece of information and the function of the text is to make random access to this specific piece of information as efficient as possible. On the most basic physical level, this kind of reference reading was facilitated by the codex format, which facilitated random access to the text not only by allowing the reader to quickly traverse the book by thumbing through it, but also by dividing the text onto discrete pages which could be numbered and referenced in indices and concordances of various kinds (Chartier 1995: 19), essentially defining a ‘textual coordinate system’ by which the book could be navigated. However, unlike modern cookbooks which often dedicate an entire page or even a whole spread to a single recipe, medieval recipe collections usually contain several recipes to a page, which means that the page is not the most useful frame of reference. Recognising this, many recipe collections—including three of the six versions of the *PD* (MSS Ad, D and H279) identify individual recipes by explicitly numbering them.<sup>50</sup> In addition to (or instead of) the numbering of recipes, all of the versions use various visual means of indicating the beginning of a new recipe and to highlight the titles (and the number) of the recipes.

Another organizational device used to facilitate reference reading is the *table of contents* which allows the reader to find the relevant item even more quickly than thumbing through the whole book (Carlquist 2004: 108). Of the six *PD* versions, all except for MS H4016 contain a table of contents, listing the recipes included in the collection in more or less the same order as they appear in the collection itself.<sup>51</sup> In addition to indicating the order of the recipes, three of the five tables also number the recipes,<sup>52</sup> effectively creating what Gunder (2001) and Carlquist (2004) have—in analogy to the digital hyperlink—called an “analogue link”, pointing to the corresponding recipe within the collection and allowing the reader to jump directly to the relevant part of the collection without needing to browse through the titles of intervening recipes. While the presence of at least some of these pragmatic text-organising devices geared towards reference reading in all of the *PD* versions strengthens the case for their function as practical reference tools

<sup>50</sup> Of the versions that do not have original recipe numbers, MSS As and C have had recipe numbers added by later (modern) users of the collection, although the numbering in MS As stops after the first 14 recipes. As Carroll (2006: 320-1) points out, the fact that these and many other originally unnumbered collections have been numbered later—by modern editors at the latest—is a telling indication of the importance of numbering as a requirement of efficient reference reading.

<sup>51</sup> There are some slight discrepancies in most of the tables, mostly involving recipes omitted either from the table or from the collection itself and recipes occurring in transposed order. For some reason, the table of contents in MS As has been left incomplete, covering only half of the collection and being followed by four empty folia which have been ruled in preparation for the table but never filled.

<sup>52</sup> The tables of contents included in MSS Ad and C do not include numbers. It is interesting that the table in MS As contains number references to the recipes even though they have not been numbered in the collection itself; it is possible that the rubricator was supposed to add marginal numbers similar to those found in MSS D and H279 but for some reason failed to do so.

instead of mere displays of status (see subsection 8.4.3), the significant variation in the degree to which these devices are employed could also be taken to indicate variation in function between the different versions, in turn implying a variety of possible functions for medieval recipe collections in general.

### 8.2.2 Colonies within colonies: manuscript miscellanies

In addition to variation in the internal structure of recipe collections, the physical manuscript context in which they appear also varies significantly. As Redon, Sabban and Serventi (1998: 2) have observed, medieval recipe collections occur in a variety of manuscript formats, including scrolls or *rotuli* (e.g. the oldest version of the *Viandier*) and independent codices of varying format and quality, but perhaps the most interesting manuscript context in which recipe collections are frequently found is the kind of composite codex known as a *miscellany* or a *commonplace book*, which generally contain a collection of texts found useful, entertaining or edifying by the compiler of the miscellany, representing a variety of different genres:

Commonplace books, it is argued, are rather haphazard, amateur productions which have been compiled over a period of according to the whims of their owners. In these miscellanies the book compiler's selection of material could be influenced by local, practical, domestic, or even political considerations, and the result is an intriguing and sometimes bizarre *mélange* where the modern reader gains some indication of the interests and habits of mind of an individual book producer. (Boffey and Thompson 1989: 292)

As Carroll (2003: 156) points out, these kinds of miscellanies or commonplace books are themselves best seen as discourse colonies, being by definition intended to be read referentially instead of sequentially (Carlquist 2004: 109). The fact that they frequently contain other discourse colonies—like recipe collections—within them makes them essentially ‘second-order’ discourse colonies, similar to the newspapers, hymn books and TV magazines discussed by Hoey (2001: 76, 87). In terms of the nine properties of discourse colonies defined by Hoey (2001), Carroll (2003: 156-8) finds miscellanies and commonplace books to exhibit seven of them, differing from recipe collections by not usually having a framing context (if one does not consider the physical book as such). Although especially the concept of a *commonplace book* is sometimes used in a restricted sense to refer only to personal autograph manuscripts written entirely in the hand of its producer and owner, the concept of commonplace or miscellany book is here used in a more general sense, referring simply to manuscripts that contain several distinct texts that have for one reason or another been collected together, regardless of their individual origins or production histories. Furthermore, the concept is here not limited to strictly personal collections, but extended to cover also those collections which are sometimes called “household miscellanies”, being compiled incrementally over a long time by a series of different scribes associated with a single household or other social institution and reflecting the practical preferences and requirements of that social institution instead of the personal reading tastes of any individual owner (294). The defining feature is thus considered to be not the production his-

tory of the miscellany, but rather the fact that its identity stems not from the thematic, formal or linguistic unity of texts included in the collection—which could cover a huge variety—but from the particular requirements and preferences of their owners (Treharne 2011: 228).

In terms of their method of production, which also affects their final form, miscellany manuscripts can be divided into two basic types: collections consciously ‘designed’ and copied by a single scribe, whether by an amateur scribe for his own use or by a professional one either speculatively or on commission, and collections consisting of separate *booklets* acquired over time from different sources or copied by the owner himself, each containing one or more texts, and subsequently put together into a book. Theoretically, these two types of collections should be easily distinguishable, the former being characterized by a single hand, uniformity of writing material throughout the manuscript, and a coherent visual design, and the second by the following characteristics, originally defined by Pamela Robinson (1980) and summarized by Hanna (1986: 107-8):<sup>53</sup>

- 1) variation in size of leaves in different parts of a manuscript;
- 2) variation in scribal hand or in page format in different parts of manuscript;
- 3) variation in style of decoration or illumination in different parts of a manuscript;
- 4) absence of catchwords at ends of quires (which may indicate once independent sections of a manuscript);
- 5) independent sets of quire signatures in different parts of a manuscript;
- 6) soiled or rubbed outer leaves of a quire;
- 7) quires formed of varying numbers of leaves in different parts of a manuscript;
- 8) variation in size of possible final quires of a textual unit—either an excessively large quire or a quire containing very few leaves so as to exactly accommodate the end of a text;
- 9) blank leaves at the end of quires, often cut away; and
- 10) short texts, added—sometimes in later hands—in originally blank spaces at the end of quires.

However, it should be noted that the method of production in itself does not reveal anything about the producer of the manuscripts, as ‘designed’ miscellanies can be produced both by professional scribes, either speculatively or by commission, and by private individuals for themselves, and miscellanies consisting of separately acquired components can mix together booklets produced commercially and ones copied by the owner himself or by an acquaintance. For example Boffey and Thompson (1989: 295) have argued that the existence of these miscellany manuscripts consisting of individual booklets copied by different scribes points toward the ‘booklet’ having served as a convenient unit of commercial production and marketing for an audience that did not want or could not afford more expensive or extensive books. As Hanna (1986: 101) has pointed out, from the per-

<sup>53</sup> Hanna himself adds three more features to the ones described by Robinson, namely 1) variation in the material from which different parts of a manuscript are made—shifts between paper and vellum, shifts (insofar as these are recognizable) among kinds or qualities of vellum, shifts among different paper stocks, 2) variation between sources from which different parts of a manuscript have been copied, and 3) variation in subject matter in different parts of a manuscript, although the last two of these can equally well apply to miscellanies produced by a single scribe at one sitting.

spective of the buyer such a commercially produced booklet represented a commodity conveniently available in a stationer's or a bookseller's shop, presenting a desired text to be purchased and joined with other texts in a miscellany, while for the bookseller, who typically produced books to order, booklets provided a way to have some popular texts in stock without the major investment of producing an entire codex.<sup>54</sup>

In addition to individual booklets, also entire miscellanies seem have been produced speculatively. For example MS Lansdowne 699, a miscellany manuscript that contains literary material as well as practical advice like a dietary and the *Stans puer ad mensam* by Lydgate (which is also found in MS Ad), "is obviously designed to cater to a wide range of tastes—religious, historical-romantic, didactic and practical", and could very well be an example of a speculative anthology, "put together in the hopes that it would find a buyer and thus aiming to be as all-inclusive of bourgeois tastes as it could be" (Sponsler 2001: 12). Some regional book-producers also seem to have produced 'homemade' miscellanies to cater for the "rather less discriminating but no less avid reading interests of pragmatic owner-producers who had an eye for material with a utilitarian appeal, as well as for texts that reflected their own particular recreational interests" (Boffey and Thompson 1989: 297). Based on the wide variety of surviving miscellanies containing entertaining and useful material, they seem to have been owned both by more discerning, wealthy readers and by people of lesser means—who often seem to have produced more modest miscellanies for their own use—the principal difference in these books being in the level of decoration and the quality of their material and execution (Boffey and Thompson 1989: 281-2).

On the basis of the features described above, the production histories of the two miscellany manuscripts containing copies of the *PD*, namely MS C and MS Ad, would seem to differ from each other: while MS C exhibits many of the features listed above for manuscripts produced as individual booklets and subsequently collected together (see subsection 9.2.6), MS Ad seems to have been written by one or two scribes and exhibits a very uniform design throughout the manuscript, suggesting that it was produced as a single entity. Whether it was produced commercially, either speculatively or on commission, or for personal use, is difficult to judge, but the neat and proficient, even decorative nature of the cursive hand used to write it and its clear and regular layout would seem to indicate professional production. In contrast, both the hands used to write the different parts of *PD* found in MS C and its layout are much less regular in appearance, making it more likely that it was copied by a less experienced or amateur scribe.

In terms of their content, the two *PD* miscellanies are very typical examples of the format.<sup>55</sup> The earlier of them, the early 15<sup>th</sup>-century MS C, contains—in

<sup>54</sup> For example in the field of medical writing, Voigts (1989) has observed that of the 178 medical manuscripts surveyed by her, 38 were assembled from originally discrete units before 1600 (35 consisting entirely of 15<sup>th</sup>-century booklets), testifying to the relatively common incidence of booklet circulation: "Booklet compilation was clearly an important element in the late medieval production of scientific and medical manuscripts, for not only did it provide for the construction of anthologies, it also resulted in codices that combine older texts with more recent commentary" (Voigts 1989: 356).

<sup>55</sup> Other examples of late-medieval miscellanies described by earlier scholars include BL MS Cotton Julius D.viii, a fifteenth-century "institutional 'household book'" in Latin and English described by Voigts (1989), which apparently comes from Barking Abbey and contains "ecclesiastical and secular chronicle material; agricultural writings; instructions for bleaching linen; recipes for making



addition to the *PD*—two other recipe collections,<sup>56</sup> several collections of medical recipes and miscellaneous medical advice, a trilingual herbal, and some religious material such as a treatise for a hermit or an anchorite, while the late-15th-century *MS Ad* contains two treatises on gardening, Lydgate's English translation of *Stans puer ad mensam*, medical and astrological texts, several histories, and John Shirley's translations of the *Livre de Bones Meurs* and the *Secretum Secretorum*.<sup>57</sup> While the contents of *MS Ad* display an emphasis on good manners and seem to reflect general genteel interests, those of *MS C* are quite different, with medical recipes and *materia medica* taking up most of the space not occupied by culinary recipes.

As Skaarup (1992: 40) has pointed out—contrary to Voigts's (1989: 348) claim—this association between culinary recipes and medical material is a frequent one, “a high proportion of medieval culinary collections” occurring in volumes “otherwise entirely devoted to medical matters” (Hieatt 1996: 54).<sup>58</sup> This co-occurrence of culinary recipes and medical treatises emphasizes the “close logical rapport” (Scully 1992: 42) between food and medicine and makes “clear how much of the medieval attitude towards food was based upon ancient theories of a healthy diet” (Strong 2002: 80). In some instances it appears that the person who had the manuscript compiled or copied—whether a physician or an aristocrat—saw no reason to distinguish between culinary and medical materials (Scully 1992: 43). This should not come as a surprise, since ultimately both medical treatises and culinary recipes were concerned with health and well-being, and their co-occurrence would seem to point towards an overlap in the readerships of medical and culinary texts.<sup>59</sup> In

---

medicines, soap, parchment and ink; instructions for cutting stones; a cookery treatise; and the charge to the cellaress of the Abbey” (389). Another example is BL MS Sloane 2027, a large paper miscellany described by (Meale 1989: 216), which was possibly owned by a late-15<sup>th</sup>-century minor provincial landowner, “Wylliam Braundon of knolle in the Counte of waryke” and contains a copy of the *Brut*, Vegetius' *De Re Militari*, John Russell's *Boke of kerving and nurtur*, and Lydgate and Burgh's *booke Off the gouernaunce off kyngis and pryncis*. Beinecke Library MS 163, a miscellany described by Hieatt (1988) and containing a copy of the *Ordinance of Pottage*, along with “such unrelated materials as a parliamentary text; medical recipes; treatises on astronomy, hunting and the interpretation of dreams; a poem on hawking; and a charm against thieves—among other things” (9).

<sup>56</sup> The first of these is a fragment of a collection which seems to contain some recipes from the *Liber utilis coquinariorum* (edited in Hieatt and Butler 1985: 81–91 from BL MS Sloane 468) and the second is a copy of the *FC* (used for collation by Hieatt and Butler 1985).

<sup>57</sup> See appendix F for a detailed description of the full contents of these manuscripts.

<sup>58</sup> According to Voigts (1989), texts having to do with “domestic endeavours” such as cooking and lacemaking “are not commonly met with in the company of ‘scientific’ writings”, but rather in “household books” (348). Whether it is the case that medical material does commonly occur in these “household books” (a term which is perhaps better suited to describing those printed books of the 16<sup>th</sup> and following centuries which are explicitly identified as such) or that cooking texts do indeed occur in medieval scientific miscellanies, is largely a matter of semantics. As for example Hieatt (1992) and Jones (2000) have observed, also individual culinary recipes—although mainly for various sugary and spicy confections and drinks, which can be considered quasi-medicinal in nature (Hieatt 1992: 16)—also frequently occur in manuscripts containing medical material, either as marginal or flyleaf additions, as in BL MS Sloane 442 and Wellcome Library MS 542, or scattered among medical and other recipes, as in Oxford All Souls College MS 81 and BL MS Harley 2378 (Jones 2000: 315). According to Jones, these kinds of individual recipes include both jottings down of personal recipes and ones copied from an established collection, but unfortunately she does not indicate whether her examples belong to the former or the latter category.

<sup>59</sup> Scholars—such as Terence Scully—who have studied medieval recipe collections on a large scale, have noted that also cookbook authors often show a profound knowledge of medieval theories of

her research on the regimen sanitatis tradition, Weiss Adamson (1995: 9) has also found evidence that medieval physicians participated both in spreading recipes and culinary practices, and in circulating culinary manuscripts.

From the point of view of the modern scholar, the inclusion of recipe collections in miscellany manuscripts have the advantage that the other texts included in the miscellany can provide us with clues about the interests of the intended target audience of the text. It should be noted that this does not always mean the owner of the manuscript—as was pointed out above, the text selection included in a miscellany might be determined by and reflect not only the individual tastes and preferences of its eventual owner, but also the requirements of the bespoke trade, the speculative activity of a bookseller, or simply the fortuitous availability of texts or even a process of gradual collection or even accidental accretion of material (Gillespie 1989: 326; Greetham 1992: 69; Machan 1994: 166). However, what the inclusion of culinary recipes in miscellanies containing material not related to the functioning of the kitchen does tell us, is that interest in medieval culinary recipes was not limited to the kitchen. Although it is quite possible that some copies of recipe collections surviving as separate volumes were used in the kitchen by a cook or his kitchen clerk, the copies included in miscellany manuscripts reflect an audience with more diverse interests than a professional cook, and it seems more likely that they were used in the study than in the kitchen.

### 8.3 The recipe as a genre and a text type

Even though Görlach (1992: 745) has claimed that the recipe text-type has “an age-old name”, Carroll (1999: 28) quite soberingly points out that in fact “the word cognate with Modern English *recipe* is not used to name a text-type before the end of the fourteenth century”,<sup>60</sup> and for example in the *PD* collection, the word does not occur even once. This is in fact not very surprising, considering that all of the examples included in the *Middle English Dictionary* (*MED*) use the word in a medical context.<sup>61</sup> Even of the printed cookbooks or culinary recipe collections of the 17<sup>th</sup> century, most never use the word *recipe* to describe their contents, but rather use expressions like “secrets”, “experiments”, “curiosities” and “ways of making”. Upon a cursory inspection, it would seem that it was only in the 18<sup>th</sup> century that the term *recipe* achieved its modern hegemony in describing instructions for culinary preparations. Although the use of the term for Middle English recipes is thus an anachronism, it is nevertheless adopted here as an analytical term for the text type in order to emphasize its historical continuity.

While the basic “culturally established tasks” (Egins and Martin 1997: 236) accomplished by recipes—namely the provision of instruction for preparing a product—may have remained the same, the different levels of context in which they are accomplished, and therefore also the means—i.e. the register—used for accomplishing them have changed considerably between the late medieval period and

---

nutrition in their choices of terminology, use of medical source material, or inclusion of dishes specifically meant for sick people (Weiss Adamson 1995: 9).

<sup>60</sup> Even then, it is used to refer exclusively to medical and alchemical recipes (Carroll 2003: 149).

<sup>61</sup> In the *OED*, the earliest culinary citation for *receipt* is from 1595, and the earliest one for *recipe* from 1743 (Carroll 2003: 149).

the present day. This means that we must be wary of relying too heavily on 21<sup>st</sup>-century genre boundaries which are “frequently inappropriate” and “misaligned with those of earlier periods”, skewing our perception of medieval texts (Treharne 2011: 229). As we have seen in this and the preceding chapter, it is not only the material context of food preparation that has changed, but also the very definition of good food, including the criteria used to judge it, have changed tremendously. This means that instead of evaluating medieval recipes and the dishes described by them from a modern viewpoint—which has frequently led to both medieval recipes and the dishes themselves being deemed unsuccessful or inadequate—we must start from the assumption that medieval recipes were a functional means of communicating information required for the preparation of a dish that was considered desirable in its cultural context. This kind of ‘bottom-up’ approach will allow us to proceed from the surviving evidence towards a reconstruction of both the contemporary cultural norms that defined ‘good food’, and the communicative norms that defined the appropriate ways of communicating the information required for fulfilling these norms.

### 8.3.1 Textual characteristics of recipes

While the genre of recipes is determined by its culturally recognized function of providing instruction on how to prepare something, it is also defined as a *text type* by the set of formal textual characteristics that are typically employed in fulfilling this purpose (Carroll 1999: 28, 2003: 178) and allow us to recognize a piece of text as a recipe:

Text-types are part of a speaker’s linguistic knowledge. A member of a linguistic community will be able to identify an example of a text-type, to name the text-type that the example belongs to (if it has a name), and to determine whether the particular text has correctly used the expected or obligatory linguistic features and formulas of the type. (Carroll 1999: 27)

However, although the generic function of recipes may have remained more or less the same since the Middle Ages, their textual characteristics have not. While modern recipes are instantly recognisable as recipes not only by their subject matter but also by their textual form, consisting of a title, a list of ingredients with quantities, and the use of imperative verbs, medieval recipes are much less marked in their textual form, and do not correspond to the modern notion of what a recipe should look like (Meredith 2004: 29). As described in chapter 2, this analysis of late medieval recipes as a text type follows Carroll (1999: 28) in adopting the view of Biber (1988 and 1995) that text-types are categories defined exclusively by their linguistic characteristics, in contrast to *genres* and *registers*, which are defined on other, functional and contextual features of the communicative situation. This does not, however mean that text-types are independent of functional or contextual factors: on the contrary, certain text-types—i.e. clusters of textual and linguistic characteristics—are frequently associated more or less intimately with—or ‘employed’ by—certain genres and registers. However, it is important to note—as pragmatics and discourse analysts have long been aware—that the mappings between pragmatic functions and linguistic forms “are more often many-to-one

than one-to-one” (Carroll 2006: 313), and specific text types should not be equated or exclusively associated with specific genres or registers and vice versa. However, because of the pervasive nature of these associations—which may have pragmatic or merely conventional motivations—text-types are frequently interpreted functionally in terms of their purposes and other situational characteristics, as Biber (1995: 10) points out. In other words, text-types—as sets of conventions governing the textual features of a textual object—are both determined by and influence several levels of the context—cultural, situational and textual—and are their elucidation is therefore central to understanding the context of medieval culinary recipes and the role they play in it.

### Textual structure of recipes

Perhaps its most studied aspect of medieval recipes as a text type is its information-structure, which has been discussed at least by Stannard (1982), Hunt (1990), Görlach (1992 and 2004, Jones (1998), Carroll (1999, 2003, 2004 and 2006), Taavitsainen (2001), Cabrera-Abreu (2002), Grund (2003) and Mäkinen (2004). The majority of these studies have divided the informational structure of recipes into anything from four to six components, some of which are generally held to be optional.<sup>62</sup> These divisions (seven versions of which are summarized by Carroll 2006: 308) are mostly based on the four basic components first identified by Stannard (1982: 60–71) as “purpose”, “requisite ingredients and equipment”, “rules of procedure” and “application and administration”, and subsequently reformulated by Görlach (1992: 746) in a more specifically culinary context as “title”, “ingredients”, “procedure” and “how to serve up”.

The problem with most of these studies is that in analyzing the informational structure of medieval recipes, they equate these informational categories with textual units within the recipe: for example Hunt (1990: 17) calls them “components” of the recipe while Görlach (1992: 17 and 2004: 125) names them “subsections”. While admitting that these divisions may be justified to some extent as an analytical aid, Carroll (2006: 309) warns us that they involve a “danger of imposing our present-day expectations on the medieval text, for example with respect to a distinction between ingredients and procedure”, which were usually not textually separated in culinary recipes until the 19<sup>th</sup> century (Hieatt 2012: 15).<sup>63</sup> Although most of the scholars who have studied the structure of medieval recipes concede that the ingredients do not need to be listed before the procedure, I agree entirely with Carroll (2006), who argues that “it would be more accurate, according to the

<sup>62</sup> While different scholars have emphasized different aspects of their textual structure, the majority of the studies divide the recipe into something like: 1) a title, usually describing or naming the intended product, 2) the required ingredients, 3) the procedure for creating the product, 4) the application of the product (optional), and 5) a formulaic phrase marking the end of the recipe.

<sup>63</sup> In medical recipes, the separate list of ingredients appeared already in the 16<sup>th</sup> century (for an early example, see e.g. Thomas Gale’s *Antidotarie*, published as a part of his 1563 work *Certain vvorokes of chirurgerie*) but to my knowledge, the first English recipe collection or cookbook to adopt the modern convention of presenting the ingredients and their quantities as a separate list before the procedures was *Mrs Beeton’s Book of Household Management*, written the young Mrs Isabella Beeton and published first in serial form in *The Englishwoman’s Domestic Magazine* over the period of 1859–1861 and then as a single volume in 1861. According to Hieatt (2012: 15), in American cookbooks this development is usually credited to the even later Fannie Farmer’s *Boston Cookery-School Cookbook* published in 1896.

palaeographic evidence, to say that for medieval recipes the ingredients do not constitute a distinct section at all" (310).<sup>64</sup> This equation most likely stems from the strong visual and structural separation of the ingredients and procedure in present-day recipes, which is easily seen as one of their defining characteristics.<sup>65</sup>

As the pervasiveness of the distinction between ingredients and procedure in modern recipes shows, the *visual paratext*—e.g. layout and typography—both reflects and influences the way in which textual structure is perceived. In her analysis of the palaeographical structuring of recipes in manuscript recipe collections, Carroll (2006: 306) comes to the conclusion that the division of any type of medieval recipes—culinary, medical or household—into more than two text-structural components—the title and the body (310)—is not justified in light of the manuscript evidence and does not seem to reflect “contemporary medieval perceptions of the structure of recipes”.<sup>66</sup> The same visual structure is observed also in all of the versions of the *Potage Dyvers*, where the title—when it exists<sup>67</sup>—is always clearly distinguished from the body, which in turn is presented as a single block without further subdivisions.

As Carroll (2006: 314) has pointed out different manuscript versions of the same text often use very different means of separating the recipe title from the body.<sup>68</sup> Perhaps the most traditional method for separating textual components in medieval manuscripts is *rubrication*, which is used in this function in all of the *PD* versions except for MS Ad. Another frequent means of distinction is placing the title on a line of its own, often inset from the edge of the text block and sometimes separated from the preceding and following lines by some extra space, which occurs in all *PD* versions except for MSS As and H279. In addition to these two basic methods, Carroll (2006: 313) lists a number of other, less common means of visually separating the title from the body of the recipe, including marginal titles, use of pigments other than red, underlining the title in ink, use of larger, more formal or more decorated script for the title, and the use of large initials or symbols (like a paraph or a double virgule) to mark the beginning of the body of the recipe. Almost all of these means are also found in the different versions of the *PD*, as

<sup>64</sup> She does, however, mention a single recipe in Glasgow MS Hunter 185 offering potential counterevidence, as pointed out by Alonso Almeida (2001: 217), but I have to agree with Carroll that Alonso Almeida's interpretation that *puncti* are used to separate different subsections of the recipe is much less likely than the alternative explanation that they are simply used to separate clauses, which they also happen to do.

<sup>65</sup> For example the diachronic approach taken by Görlach (1992) clearly leads him to disregard historical changes in the text type and to forcibly generalize features of modern recipes to those of earlier periods, obscuring some of the characteristic features of medieval recipes.

<sup>66</sup> However, Carroll (2006: 310-1) does mention the recipe number as a possible candidate for a third component of a recipe in addition to the body and the title, although it could also be viewed as an component or extension of the title, depending on its placement. As an example she mentions the version of the *FC* recipe collection contained in New York Pierpont Morgan Library MS Bühler 36, which has marginal recipe numbers next to the recipe title, similarly to the MSS D and H279 versions of the *Potage Dyvers*.

<sup>67</sup> Most of the recipes in the MS C version were for some reason left without titles by the original scribe (who only titled the first 12 recipes), the majority of them—but not all—having been provided titles by a 15<sup>th</sup>-century corrector or annotator

<sup>68</sup> As an example, Carroll (2006: 314) mentions the BL MS Harley 1605/3 and MS Additional 5016 versions of *FC*, noting that while the former has rubricated titles placed on the same line as the end of the body of the preceding recipe, the latter has a blank line between the end of each recipe and the rubricated title of the next one.

can clearly be seen in the diplomatic presentations of the manuscripts included in appendices B and D.

Based on the evidence of the *PD* MS texts, I must agree with Carroll (2006: 323) that as useful as it is to distinguish between different kinds of information presented in a recipe, this does not necessarily mean that corresponding ‘sections’ exist on the level of textual structure: while the title can clearly be seen to have been distinct from the body of the recipe, distinguishing between an ‘ingredients’ section and a ‘procedure’ section, even with the caveat that “these sections may be combined”, is simply a misrepresentation. Furthermore, it should be noted that even on the level of information structure, some of the components considered ‘mandatory’ may be absent. For example Carroll (2003: 146) has observed that the recipes for laces, which are otherwise quite similar to other types of recipes, do not always specify the required ingredients or include information on the application of the product. Similarly, (Meredith 2004: 29) points out that in the *Diuersa Cibaria* recipe collection, the first 32 recipes consist simply of lists of ingredients, completely omitting the procedure, while the majority of the other 30 recipes follow the usual format.<sup>69</sup>

One textual feature that does seem to be a rather consistent—and generally acknowledged—structural component of Middle English culinary recipes despite not being explicitly marked is the formulaic ending, usually constituted by a variant of “& serve it forth” (Carroll 1999: 33).<sup>70</sup> This kind of a closing formula—which can be seen to parallel the *efficacy phrase* (Jones 1998) often found in medical recipes—has little information content and is commonly understood (see e.g. Carroll 1999, Cabrera-Abreu 2002 and Grund 2003) to have primarily a text-organising function, indicating the end of the recipe. This hypothesis would seem to be supported by the fact that in some versions of the *Potage Dyvers* it has been abbreviated to the form “s<sup>a</sup>”, which essentially serves the function of a final punctuation character. In conclusion, if we look at the textual structure of medieval recipes without preconceived notions based on modern recipes, their textual structure is in fact quite simple and highly standardized, consisting of:

- 1) a visually distinct descriptive identifier (title and an optional number),
- 2) a description of how to prepare the item (with varying amount of information left implicit), and
- 3) a closing formula (consisting of text and/or punctuation).

### Linguistic features of recipes

Upon his first encounter with medieval culinary recipes in the 1960s, Peter Meredith reminisces having been struck by “the personal violence of their language”, an impression created mainly by the copious use of verbs nowadays associated mainly with violence combined with personal pronouns used to refer to ingredients (Meredith 2004: 28).<sup>71</sup> His observation that the language of Middle English

<sup>69</sup> This should not be surprising, considering that this—although usually complete with quantities—is the standard form of the recipes used in a modern professional kitchen.

<sup>70</sup> Sometimes it also occurs in a longer form like “serve it forth for a good pottage” which bears an even closer resemblance to the efficacy phrases of medical recipes.

<sup>71</sup> This impression was most likely largely due to the prevalence of cutting terms like *hack* and *hew*, which were extremely common in the 14<sup>th</sup> and 15<sup>th</sup> centuries but disappeared from use in culinary

culinary recipes is characterized by “bluntness” and “lack of frills” (Meredith 2004: 31), while not exactly analytical, is apt enough and agrees with Carroll’s (2003: 147) more analytical summary of the typical linguistic features of Middle English recipes:

- headings consist of infinitive verbs or noun phrases;
- determiners of noun phrases in the recipe body may be either possessive pronouns or articles and are not elided;
- object deletion is rare, although permitted;
- verbs are mainly in the imperative, although both subjunctive and indicative forms and *shall* constructions are also found;
- sentence structure is simple and mainly paratactic; and
- textual structure is temporally organized, often reinforced by temporal discourse markers like *then*.

Of the types of recipe heading mentioned above, the noun phrase and the verbal infinitive, the nominal form would seem to be more common in the 15<sup>th</sup> century—and is used exclusively in all the *PD* versions—while the infinitival form would seem to have been favoured by many 14th-century collections on the one hand and by the early 16<sup>th</sup>-century printed cookbooks on the other, although both seem to occur throughout the period.<sup>72</sup> Although Görlach (1992) has claimed Middle English recipes to be written in an incomplete “memorandum form” (746), Carroll (1999: 29–30) has observed that all the recipes examined by her employ complete sentences almost exclusively, which is also true of the *PD* recipes.<sup>73</sup> In terms of the verbal forms used in recipes, both Görlach (1992: 748) and Carroll (1999: 30) concur that imperative is the predominant form, with *shall* and the subjunctive occurring as minority forms, while the indicative is rare, but does nevertheless

---

contexts by the 16<sup>th</sup> century (Marttila 2009: 115–6). The choice of pronoun in referring to culinary ingredients could in turn be affected by the “vestiges which survived of the old grammatical gender” of Old English or by the grammatical gender of Anglo-Norman French (Meredith 2004: 28), explaining the use of gendered pronouns for inanimate objects. The fact that the possessive and object cases of the Middle English pronoun *hit* (‘it’) were *his* and *him* further adds to the sense of personal reference for modern readers.

<sup>72</sup> The infinitival form is used almost exclusively in the three earliest ones of the four 14<sup>th</sup>-century collections edited by Hieatt and Butler (1985) (*DC*, *DS*, and *UC*), while the latest and largest of them (*Forme of Cury*), as well as *PD* and the late-15<sup>th</sup>-century collection of Corpus Christi College MS F291 (Hieatt 2012), exclusively use the nominal form. Both forms occur side by side in the late-15<sup>th</sup>-century collection known as *Stere Htt Wele*, surviving in MS Pepys 1047 in Samuel Pepys’ library (Hodgett 1972) and in the 15th-century collection preserved in BL MS Harley 5401 (Hieatt 1996), as well as in the majority of the collections edited or described in Hieatt (2008), with the exception of the longer collections in Bodleian MS Rawlinson D 1222 (mid-15<sup>th</sup> c.), BL MS Sloane 1108 (15<sup>th</sup> c.), New York Public Library MS Whitney 1 (early 15<sup>th</sup>-c.), and Trinity College Cambridge MS 0.1.13 (15<sup>th</sup> to 17<sup>th</sup> c.), which use the nominal form exclusively. Of the surviving vernacular pre-14<sup>th</sup>-century collections from England or Northern Europe, the two Anglo-Norman collections edited by Hieatt and Jones (1986) exclusively use the nominal form, while the late-13<sup>th</sup>-century Danish collection, *Libellus de arte coquinaria*, mostly uses verbal recipe titles beginning with *wyltu maken*.

<sup>73</sup> One specific feature which is frequent in modern culinary recipes but rare in Middle English ones is the deletion of verb objects (Carroll 1999: 31). While Culy (1996: 96) found more than half of the verbs in Modern English recipes to have null objects, his analysis of a small sample of Middle English recipes indicated only 4.3 percent of verbs occurring with null objects. Görlach’s (1992: 749) study of the recipes edited in Hieatt and Butler (1985) gave similar results, indicating that over 90 percent of transitive verbs had objects.

occur.<sup>74</sup>

As to their sentence structure, Middle English recipes tend to be “simple in structure, with a high level of parataxis” (Carroll 1999: 32), which is not surprising considering that English as a language is essentially paratactic in its organization, with hypotaxis being associated with the emulation of Latin and “a more literary and hence a more estimable type of style” (Blake 1992b: 19). In terms of their temporal organization, Middle English recipes do not show great differences from Modern English recipes, instructions being normally given in the order in which they are to be carried out and the temporal structure occasionally reinforced by the *then* Carroll (1999: 31), although there are instances where a procedure—often a subsidiary one like preparing a filling or sauce—is provided out of sequence.

### Aspects of recipe titles

The names by which prepared dishes are known frequently have only an obscure history. Even today the origins, the original meanings, of the names of the English *Mawmenee* and *Sorengue* are largely a matter of speculation.<sup>75</sup> (Scully 1992: 211)

Since recipes are most commonly listed and identified by their titles or names, they have also been “an important preoccupation of those doing research in the field” (Hieatt 1996: 55). In her analysis of the linguistic features of medieval recipes, Carroll (1999: 29) observed that most of the recipes she examined were accompanied with a title or heading, which, as pointed out above, normally consists of either a descriptive noun phrase or such a noun phrase prefaced with a variant of the phrase *To make*.<sup>76</sup>

The basic function of a title in this context is an identifying label that provides a means of not only visually locating and identifying it in the manuscript context, but also of referring to it from within other discourses. In accordance with this, Scully (1992: 198) sees the original purpose of recipe titles to have been the facilitation of communication between the lord, the cook and the rest of the household: by labelling a culinary preparation with a name, the cook enabled unambiguous reference to it. It must however be noted, that not all collections assign titles to the recipes: for example the vast majority of the recipes in the MS C version of *Potage Dyvers* were for some reason not given titles by the original scribe, but were only

<sup>74</sup> Based on the description of the linguistic features of medical recipes provided by Taavitsainen and Pahta (1995: 521; see also Taavitsainen 2001: 100, 107), medieval culinary recipes bear a strong resemblance to medical recipes found in popular remedybooks—which also tend to favour imperative verb forms and second-person pronouns in action-demanding sentences—in contrast to recipes found in learned texts which tend towards passive constructions.

<sup>75</sup> Although, as Scully himself points out in a footnote, the most likely origin of the term *Mawmenee* and its cognates is the Arabic term *ma'muniyyat*, as first suggested by Rodinson (1962).

<sup>76</sup> Carroll actually described recipes “beginning” with a title, but since the title or heading of a recipe is palaeographically quite clearly separated from the recipe body itself and is not always the visually first element of the recipe (as Carroll in her later article (2006) points out), they are more appropriately described as being ‘identified’ or ‘accompanied’ by a title rather than *beginning* with it.



titled (sometimes quite mistakenly) by a later corrector or annotator (see subsection 9.2.6).<sup>77</sup> Judging from the fact that the recipes that *do* have original titles are the 12 first ones, this is more likely to be an accidental omission than a conscious decision, although it most likely would have affected the usability of the manuscript, which may be precisely why the later 15<sup>th</sup>-century annotator titled them at the same time he made corrections to the texts themselves.<sup>78</sup>

In terms of the semantic reference of the titles, most medieval dishes seem to have been named according to some physical or procedural property of the dish: its principal ingredient, its colour, its method of cooking, the type or recipe ‘genre’ it belonged to, or a combination of these (Grieco 1992: 32; Scully 1992: 198). However, although the title of a recipe usually reveals something about the recipe itself, it should never be taken at face value, as many titles—whether because of transmission errors or the ignorance of the original scribe—occasionally have little to do with the contents of the recipe, even if they at first glance would seem reasonable. As an example of this, Hieatt (2012: 16) provides the titles “Rissoles of Fish” and “Lampray Hay”, of which the first is associated with a recipe not containing fish, and the second with one that has nothing to do with either lampreys or hay.<sup>79</sup>

In accordance with their labeling function, Meredith (2004) sees the titles of recipes to have been their most conservative—and in the case of titles borrowed from other languages also exotic—part, remaining “largely unaltered for centuries, except by confusion and misunderstanding” (39). This resistance to change might also explain why there exist—in addition to recipes whose titles have been corrupted through scribal misunderstanding—a number of recipe titles or ‘families’ of titles that occur in very similar forms in several different regions and languages, but denote dishes with very few similarities (Redon, Sabban and Serventi 1998: 3). Perhaps the best known of these ubiquitous titles are the ‘white dish’ (which is often not white at all) named *blancmange*, *blanmangier*, *manjar braquo* or *blamensir*, and the characteristically paste-like *morterele*, *mortarolo*, *martarolum* or *mortrowes*. This phenomenon, which often frustrates the attempts of scholars “to track down the family trees of these terminological groups” (Redon, Sabban and Serventi 1998: 30), seems to apply particularly to English manuscripts, where recipes with identical names often turn out to describe quite different dishes (Hieatt 1992: 19). In some of these cases, it might be that a seemingly inappropriate title describes an earlier form of the dish which has since evolved into a form that no longer answers to the description of the title, although the widespread use of these common names would also tend to encourage their use for new recipes though analogy.

The converse of this phenomenon is also a frequent occurrence: there are many instances where essentially the same recipe appears under different names in different collections (Hieatt 1992: 19). This type of variation can partly be explained

<sup>77</sup> The Mazarin copy of the Viandier is another example of such a manuscript: according to Scully (1988: 8), the scribe has omitted titles for the majority of the recipes in the latter part of the collection.

<sup>78</sup> Although the fact that he did not follow the titles provided for the dishes in the table of contents would seem to indicate that he was either not very familiar with or not concerned with maintaining a link to the established culinary tradition, since the fact that the titles added by him often bear no resemblance to the established names of the dishes makes them difficult to recognize.

<sup>79</sup> As mentioned above, this problem is especially acute in the case of the MS C version of the *PD*, where the titles added by the later annotator seem to be based mostly on his impression of the first few lines of the recipe, which sometimes leads to rather strange results.

by the fact that most types of dishes naturally evolved a variety of names due to the existence of a large number of synonymous terms for types of dishes originating in different languages. Pies could be *tarts*, *tortes*, *tartlettes* or *tourteletes*; porridges could be *puls*, *paniccia* or *panica*; custards could be *crustad*, *custane*, *dariol*, *erbolat* or *letelorye*; pancakes could be *forces*, *basteln*, *cryspis*, *froyse*, *rorkoken* or *struven*; while ‘doughnuts’ of deep-fried batter could be *bingnete*, *otras*, *mistembec*, *samar-tard*, *krapfen* or *krepflin*, depending on the locality where the recipe happened to originate, or on slight variations in the preparation (Scully 1992: 201). The uncertain meaning and etymology of many of these terms often makes it difficult to recognize connections between variant titles and even to recognize ‘genuine’ recipe names from scribal corruptions. Of course, another mechanism for the proliferation of titles for a single dish—demonstrated by MS C—is scribal innovation, either in the absence of an authoritative exemplar or as an attempt to improve upon a title perceived as corrupted. The advent of printing in the late 15<sup>th</sup> century, however, changed the situation in this regard, since the fact that printed recipe collections labelled their recipes more consistently than manuscript ones served to fix the previously fluid identities and ingredients of recipes more firmly (Mennell 1986: 67).

Another feature of Middle English recipe titles that might be a result of their conservatism is the fact that many of them display strong traces of French or Anglo-Norman influence, even in late 15<sup>th</sup>-century collections such as the members of the *PD* family which are otherwise thoroughly English.<sup>80</sup> Considering the linguistic history of medieval England, it is not surprising that a great proportion of English recipes originally had foreign names—mostly French or Anglo-Norman, but also Italian or Arabic in some cases. However, the extended survival of these titles far beyond the survival of French influences elsewhere in the recipes, even when they have been corrupted in transmission to become all but indecipherable, is quite interesting, especially since these recipes often describe variations quite distinct from their continental namesakes, and in some cases even uniquely English dishes.<sup>81</sup>

### 8.3.2 Functional aspects of recipe as a *genre*

[W]e recognize texts to be intentional: they are *for* something (even if we are not sure what it is), just as they elicit our attention to the precise terms of their presentation. [...] In cases where the intention of the text is overt, as in a manual for the repair of small engines or a textbook on medical physiology, the text itself ordinarily attracts little

<sup>80</sup> These traces include the use of the French preposition *de*, the frequent use of postmodifying adjectives which are inflected according to the number of the noun, and the use of French lexical items that are not witnessed in English outside the culinary context.

<sup>81</sup> This does not seem to be a phenomenon unique to recipes, as Treharne (2011) has found the vast majority of texts in some of the late-13<sup>th</sup> and early-14<sup>th</sup>-century miscellany manuscripts (such as Bodleian MS Digby 86 and Jesus College MS 29) to have French rubrics, regardless of whether the texts themselves were in French or not. Although she does not have any answers as to the reason for this, she does suggest several alternatives: “was this because the texts were known by French titles, or was French the language of finding one’s way around a large collection like this? Or was French the prestigious way of labeling a text” (232)?

attention, because it appears to us subsumed in the subject it treats.  
(Searle 2004: 4)

While it is the set of formal characteristics typical to culinary recipes, discussed above, that defines them as a *text type*, they are also defined as a *genre* by their culturally defined function, which Görlach (1992) has defined quite simply as the provision of “instruction on how to prepare a meal” (745). In addition to Görlach, also Swales (1990) has considered culinary recipes to be a genre whose purpose is especially easy to identify. According to him, they would appear to be “straightforward instructional texts designed to ensure that if a series of activities is carried out according to the prescriptions offered, a successful gastronomic outcome will be achieved” (Swales 1990: 46). Furthermore, Taavitsainen (2001: 89), observes that recipes have maintained this basic function throughout the history of English and can therefore be readily identified, even among historical texts.

The functional definition of the general term *recipe* in modern research literature has varied between different scholars of medieval texts, as Carroll (2003: 145) points out. Some, like Braswell (1984), Görlach (1992, 2004) and Taavitsainen (2001) restrict it to culinary or medical instruction, excluding other kinds of practical sets of instructions, but others, like Keller (1971), Hargreaves (1981), Stannard (1982) and Carroll (2003) use it in a wider sense to refer to various kinds of short texts providing “instructions on how to prepare something” (Carroll 2003: 146). This thesis takes the latter view, defining recipes as a *genre* characterized by the culturally defined function of *providing practical instruction for the process of preparing some physical product*. While I share the belief of (Scully 1992: 255) and others, embodied in the above definition, that medieval recipes are not mere documentation of conspicuous consumption but primarily practical tools, they do also have a significant documentative function, being essentially selective descriptions of action:

A recipe artificially isolates the actions and ingredients needed to prepare a single dish. In a real kitchen, many dishes are being prepared at the same time, and work processes and ingredients for them overlap. A recipe is a cross-section of a portion of the work going forward in the kitchen.  
(Wheaton 1983: xix)

This view of recipes as descriptions of action is the one feature that seems to be common to all definitions, and is especially obvious in medieval recipes, where the textual content of a recipe consists entirely of a description of the procedure for creating the product, any mentions of ingredients or implements being subsumed within it (Carroll 2003: 146). This action-oriented nature of recipes is also emphasized by their temporal organization on the basis of “experiential iconicity”, often reinforced with temporal adverbs (Taavitsainen 2001: 98). This association with physical action means recipe texts can be seen to resemble dramatic texts in the sense that in neither is the text itself the ‘original’, as Simon (2000: 145) has pointed out in the context of medieval plays. These kinds of texts are simultaneously records of a pre-existing physical act, and instructions for the recreation of this act, and could thus be considered instrumental or utilitarian in their nature, rather than significant in themselves. This utilitarian nature of both play

and recipe texts is also illustrated by the observation that the majority of the documents containing them seem to have been considered as *objects to be used*, being “written in secretary hands or chancery cursives, not the stately bookhands patrons would demand” and “sparingly rubricated” and “rarely illustrated” (Simon 2000: 145-6).

As was mentioned above, this thesis does not consider culinary recipes—as opposed to other types of recipes such as medical and household ones—as a genre of their own, but rather as a relatively broad *register* within a more general genre of *recipes*—which is defined by the purpose of providing instructions for producing something—distinguished by being related to the *field* of food preparation.<sup>82</sup> This kind of an inclusive definition of the recipe genre is supported by the fact that in the Middle Ages medical and culinary recipes were paralleled by “codicological and agricultural recipes”, as well as “instructions for creating paint colours” and other household supplies, which not only have the same culturally defined function of providing instructions for the preparation of a physical item, but also “demonstrate the linguistic characteristics of the Middle English recipe”, suggesting that not only the genre of recipes, but also the associated text-type “encompassed even more than culinary and medicinal recipes” Carroll (1999: 38).<sup>83</sup> This formal and functional resemblance is further strengthened by the fact that medical, culinary and household recipes often occur in the same manuscript collection Carroll (1999: 37).<sup>84</sup> Thus Carroll (2003) considers the various kinds of instructional texts to share so many features among themselves—regardless of their specific field—that “it is highly plausible that for medieval producers and recipients they were tokens of the same (unnamed) text-type” (149).

In discussing the concept of *restricted registers*, Mathiessen (1993: 254) has used culinary recipes as an example of such a restricted register. According to him, “the semantics of culinary instruction deployed in written recipes” are “quite a simple semantic variety”, the register of culinary recipes being thus fairly restricted:

[...] the field is one of procedures in the culinary realm. The mode is written and instructional. Interpersonally, the writer can either choose to interact with the reader by instructing or informing him/her or just choose to qualify some instruction. Ideationally, the writer represents either a culinary doing or a culinary being, with states of wanting/liking as a third minor option. (Mathiessen 1993: 254)

### Recipes as *Fachliteratur*

In a wider context, recipes have often been considered a form of *Fachliteratur*. This concept, defined in German textual studies, includes all kinds of practical writing,

<sup>82</sup> The further specification of a particular *tenor* and *mode* would correspondingly result in the specification of narrower registers within this broader one, such as ‘professional culinary recipes’, ‘culinary recipes for beginners’ or ‘culinary recipes presented in television’.

<sup>83</sup> For example the *Directions for laces*, a medieval instruction manual for making braided laces, was found by Carroll (2003: 148-9) to share both the textual organization, general function and linguistic characteristics of culinary, medicinal, and codicological recipes, and to thus be representatives of the not only the same *recipe* genre, but also of the same text type.

<sup>84</sup> Also many of the early modern printed cookery books include sections on “preserving, dairying, brewing and distilling”, as well as “a substantial number of medical recipes”, being thus “household books in the fullest sense” (Wilson 2004: 24).

excluding only devotional and belletristic writing, corresponding better to medieval classifications of knowledge than stricter definitions of ‘scientific’ writing that have been more widely used in Anglo-American scholarship (Voigts 1989: 347).<sup>85</sup> Topics that are included in various medieval definitions of practical science and thus of *Fachliteratur* include architecture, agriculture, alchemy, armament, augury, commerce, cooking, fabric-making, hunting, navigation, necromancy, science of mirrors, tailoring, and theatrics (347). This concept of *Fachliteratur* has been applied to the study of medieval texts for example by the essays included in Eamon (1982) (of which Stannard’s specifically discusses recipes as *Fachliteratur*) and by Laurel Braswell (1984), who takes a particularly inclusive approach, including in her discussion not only what we would now call ‘scientific’ texts and “those which made up the quadrivium and occult and prognostic writings”, but also “manuals for conduct and for chivalric, military, and sporting endeavours; grammar books; recipe books for cooking and book production; guidebooks for travel; and miscellaneous treatises on such subjects as sailing, lacemaking, and the assize of bread and ale” (Voigts 1989: 347).

The concept of *Fachliteratur* has also influenced the definition of medieval *scientific* writing by Voigts (1989: 348), whose definition of this category is based on its subject matter and is more inclusive than the definitions used by earlier historians of science but less inclusive than that of *Fachliteratur*. Voigts delimits the topics covered by medieval scientific writing on the basis of a threefold division by Manzalaoui (1974) into 1) “activities that are experimentally sound, mathematically true, or empirically useful (geometry, astronomy, pharmacology, herbal lore)”; 2) “pseudo-sciences or consistent logical systems involving study but which cannot now be substantiated by experimental fact (dream lore, lapidaries, judicial, astrology, physiognomy); and ” 3) “the occult (alchemy, geomancy, chiromancy)”, supplementing this division with “other kinds of medical writing and agricultural treatises” (348). This means that while culinary recipe collections are included in the sphere of *Fachliteratur*, they are not usually considered to belong to the category of medieval scientific writing even according to its loosest definition, unlike the closely related register of medical recipes.

Another classification that sounds misleadingly similar but is in reality very different, has been defined by Butcher (2004), who divides medieval texts into *literary texts* and *pragmatic texts*. However, his definition of pragmatic texts includes only administrative records and business documents, relegating *Fachliteratur* into the category of literary texts, creating a rather unbalanced categorization. While misleading in its choice of terminology, his classification can nevertheless be combined with the concept of *Fachliteratur* to offer a more or less complete coverage of the medieval textual universe under a threefold division into:

- 1) *instrumental texts*, covering the kinds of administrative records and business documents described by Butcher, as well as legal writs and official proclamations that serve as instruments of administrative or legal actions (with performative force), or records of information required by them;

<sup>85</sup> There is evidence that some scribes were specialising in the copying of texts that could be characterized as *Fachliteratur* already in the early 15<sup>th</sup> century (Voigts 1989: 384), an example being the scribe designated as *Delta* in Parkes and Doyle (1978: 206-8) and associated with five manuscripts, of which all are prose and four of which represent *Fachliteratur*.

- 2) *utilitarian texts*, covering *Fachliteratur* in its inclusive form, as defined for example by Braswell (1984), as well as almanacs, lunaries, and various other kinds of texts whose function is to provide information required for performing practical operations; and
- 3) *literary texts*, covering not only imaginative or belletristic works but also all historical, religious and philosophical works that are intended to entertain or edify their readers, or to convey knowledge *an sich*, without regard to practical utility.

This kind of high-level classification into what could be considered ‘genre types’ may not be directly useful in analysing the functions and features of individual texts, but by serving as a basis for a functional textual taxonomy, it can help us relate different genres and registers to each other in a meaningful way. In this kind of a taxonomy, we could for example postulate the genre of *recipes* as a descendant of the *utilitarian texts* genre type, a sibling of the genres of *manuals*, providing instructions on how to perform a procedure and *guides*, providing instructions on how to achieve a state of affairs of some kind. These two genres, in turn could be seen as children of the genre subtype (or ‘macro genre’) of *instructions*, providing information on *how* to do things, which could in turn be seen as paralleled by the genre subtypes of *maps & charts*, providing information on *where* to find things, *calendars*, providing information on *when* to do things, and *diagnostics*, providing instructions on how to recognize *what* a thing is, as sketched in Figure 8.1.<sup>86</sup> These subtypes, answering the questions of *how*, *what*, and *when*, respectively, could in turn be assigned to an intervening category if required, in turn parallel to other, more distantly related types of utilitarian texts, just as the genres themselves can be subdivided into increasingly specific registers based on the three dimensions of the situational context defined by Halliday (1978) and outlined in subsection 2.2.1 (see footnote 27 on page 34).

### 8.3.3 Vagueness of medieval recipes

One topic of frequent commentary about medieval culinary recipes is their information content, which—as was pointed out by Carroll (2009: 57)—has been characterized quite differently by different scholars. While Hieatt and Butler (1985), medievalists familiar with culinary recipes, characterize them as “precise and discriminating in directions for seasoning and colouring” although somewhat “terse” (8), Görlach (2004), a historical linguist, and Lehmann (2003), a social historian, merely see them as “imprecise” (Görlach 2004: 125) and lacking “what we would consider workable instructions” (Lehmann 2003: 20). What the use of such derogatory characterizations as “imprecise” and “absence of proper quantifications” (746) by Görlach and Lehmann reveals, is that they are in fact judging medieval recipes in terms of their own modern conceptions of what a recipe should be like. This

<sup>86</sup> Practical examples of these genres and genre subtypes—taken from the field of medicine—could be *regimina sanitatis* describing how to maintain a state of good health for the genre of *guides*; practical surgical treatises describing how to perform surgical operations for the genre of *manuals*; herbals, lapidaries and diagnostic lists of diseases for the genre subtype of *diagnostics*; prognostications, lunaries, and almanacs, providing information on the timing of actions for the genre subtype of *calendars*; and descriptive anatomies for the genre subtype of *maps & charts* (although most works of anatomy also contain parts representing other genres).

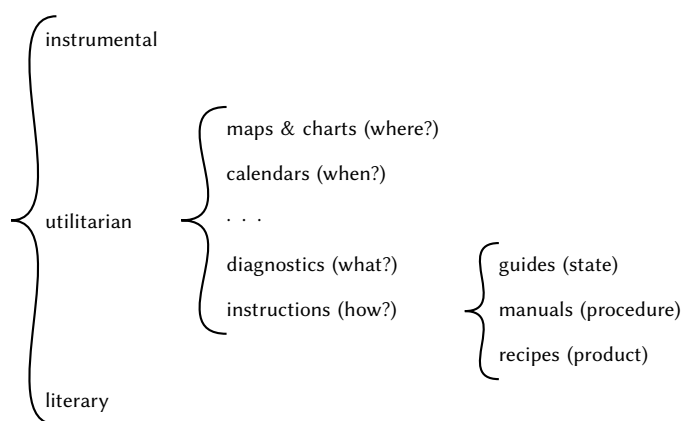


Figure 8.1: A sketch of a possible functional taxonomy of genres.

modern perception of the vagueness of medieval recipes is further sharpened by our modern view of instructional texts as “one of the contexts in which specificity is required and vague language is inappropriate” (Carroll 2009: 56). These expectations and the failure of medieval recipes to fulfil them supports the view that while text-types are defined on linguistic grounds, they are constantly changing in culturally determined ways, and that the customary situational context (especially the relationship between the author and user) of medieval and modern recipes are likely to be significantly different from each other.

While the most common explicit linguistic expression of vagueness discovered by Carroll (2009: 65–6) is *non-numerical vague quantification*,<sup>87</sup> this is still in clear minority when compared with the complete *omission* of quantification, which is a much more significant contributor to the perceived vagueness than any linguistic expressions. One of the reasons for the seemingly conflicting statements about the level of detail in Middle English recipes is that while they often omit information like quantities of ingredients or cooking times, they may nevertheless contain very precise directions for the seasoning and colouring of a dish—aspects of the dish that are often seen as ones best left to individual taste in modern recipes (Hieatt and Butler 1985: 8).

Why, then, are medieval recipes so different from our modern expectations? To a large extent, this difference may in fact be largely illusory and based on our mistaken assumption that “all recipes are written for the same function and for the same audience” (Carroll 2009: 57). The differences that are usually seen as existing between medieval and modern recipes, may in fact be differences between recipes intended for the everyman and recipes intended for culinary professionals (Scully 1992: 24).<sup>88</sup> Thus the apparent lack of essential information in medieval recipes can be seen as a result of the different amount of *common ground* (Lee 2001: 41)

<sup>87</sup> Using such quantifiers as ‘a little’, ‘a few’, ‘a part’, ‘a portion’, ‘a quantity of’, ‘some’, ‘enough’, ‘a good quantity’ or ‘a great quantity’ which occurs 65 times in the 205 edited recipes of the *FC* collection (Hieatt and Butler 1985).

<sup>88</sup> Even today, recipes written down as personal *aides-mémoires* or for use by colleagues by a professional chef, are likely to omit various pieces of ‘required’ information much in the same way as medieval ones do.

shared between the author of the recipe and its original audience on the one hand and its modern readers on the other. The different roles of the cook and cookery in society, along with differences in the role of written texts in general and in the extent of literacy mean that the medieval discourse communities using and producing recipes were likely to be very different—and more restricted—from modern ones, resulting also in “very different expectations for recipe texts” (Carroll 2009: 77). Regardless of whether their original audience was an “expert cook, a lord or lady dictating a menu, or someone aspiring to the status and wealth inherent in such use of fresh meat and exotic spicery”, he or she was at least expected to share the same cultural context with the author and be “familiar with the dishes being described” through prior experience (Carroll 2009: 79). If we furthermore assume the recipes to be intended not for complete kitchen novices (as modern printed cookbooks generally are) but for people with practical experience of cooking—or even seasoned professionals—these readers would have possessed much of the ‘missing’ information implicitly, being able to deduce the appropriate quantities, temperatures, and cooking times based on their experience and empirical observation of the dish during the cooking process.

This kind of dynamic and adaptive approach to cooking is also supported by the presence of various indirect means of indicating the appropriate cooking times and ingredient quantities, such as the use of adverbial clauses containing an adjective describing the intended result (e.g. “add flour *until it be stiff enough*”).<sup>89</sup> However, since the modern reader—unlike our medieval forebears—does not know what the dishes were supposed to look and taste like, “we are left to guess temperatures or cooking times, or how much of a given ingredient to add”. (Weiss Adamson 2004: xviii). Also the physical realities of medieval cookery, namely the difficulty of maintaining constant temperatures and the impossibility of recording and communicating them, mean that the concept of absolutely measured cooking temperatures is simply irrelevant; while instructions such as “cook it on a gentle fire,” or “make a tiny fire,” (Weiss Adamson 2004: 63) might seem vague to us, these kinds of relative instructions are in fact as accurate as it is possible to provide when cooking on open fire without the aid of a thermometer (Scully 1992).

Similarly, in the absence of accurate clocks and standard measures of time, it is obvious that a recipe “never provides cooking times in hours or minutes”, and “a comment referring to a generally known activity like saying a prayer or walking a certain distance” (63)—despite its inherent inaccuracy—is in fact one of the only reasonable ways of specifying time in cases where it is judged to be essential. It was part of the cook’s professional skill “to know from experience—or from intuition—what temperature was appropriate for a certain dish, or for a certain step in the preparation of a dish” (63) or to judge when a dish had been cooked enough. Similarly, in a situation where the potency of spices and could vary considerably depending not only on their exact species and origin, but also on how long they had been kept and in what conditions, and how much they had possibly been adulterated by traders on their long way to the kitchen. Similarly, even the strength of flour and the exact properties of other foodstuffs could vary, making it a far more reasonable approach to use one’s experience and empirical

<sup>89</sup> This kind of indirect indication of quantities and times is in fact one of the principal functions of adverbial clauses containing adjectives (e.g. ‘until it be [ADJ] enough’, ‘to make it [ADJ]’, ‘so that it be [ADJ]’) in Middle English culinary recipes.



observation, together with descriptive phrases such as ‘until it be stiff enough’ or ‘to make it poignant’ to judge the appropriate quantities than to rely on absolute measurements. This means that much of what we consider as indispensable information in modern recipes, would have been either obvious or simply impractical (or even impossible) for the medieval reader.

The fact that absolute quantities were much more impractical in the medieval context than today does not mean that they are entirely absent from the recipes:

In addition to the occasional gallons, quarts, pints, pounds, ounces, inches and the like, quantities and sizes are often expressed with the help of other foodstuffs, such as eggs, or nuts, or parts of the body, such as the length and width of a finger. And then, of course, there are the relative measurements ‘twice as much as,’ ‘a quarter of the amount of,’ or simply ‘not too much of’.

(Weiss Adamson 2004: 63-64)

As Carroll (2009: 64) points out, this kind of explicit quantification is especially frequently used to indicate *size*, either by reference to other common objects, using phrases like “of a mannys lengthe” (MS Ad f.56r), “iij fyngurs brede” (MS Ad f.38r), “of a peny brede” (MS As f.17r), “þe brede of a pewtre disshe” (MS As f.37r), or using quantified measurements like “an enche of hegh” (MS H4016 f.6v). There also seem to be significant differences between individual collections. In addition to the unusually detailed late-15<sup>th</sup>-century Corpus Christi College MS F 291, where “an unusually large number of the recipes [...] give exact quantities for at least some of the ingredients”, Hieatt (2012: 13) has noted only three other collections that provide some quantities, although nowhere near to the same extent as the Corpus Christi manuscript.<sup>90</sup> Carroll (2009: 63) has furthermore noted that when quantities are provided, they are often for alcoholic beverages or sweeteners, which are ingredients often used in medicinal preparations. Since medical recipes do usually provide quantification, it might well be that the quantification of these ingredients even in culinary recipes is due to their association with medical uses. The fact that some recipes *do*—albeit rarely—quantify at least some ingredients also leads (Carroll 2009) to the conclusion that the lack of measurements in the majority of recipes is “due to different textual norms and expectations rather than a lack of means or competence” (62-3).

One aspect of medieval recipes where the lack of explicit quantification has caused difficulties and controversy for modern scholars is their degree of spicing (Lehmann 2003: 21). Some scholars, like Hieatt (2012), have considered the spiciness of medieval food to have been overestimated, interpreting the lack of quantities in the recipes as an indication that “the quantity was so small as to be unremarkable” (20), while others like Bruno Laurioux, have taken the large number of different spices used in medieval recipes as an indication that medieval food was more spiced than what we are used to (Laurioux 2002: 19–21). The evidence found by Hieatt (2012: 13-4) in the Corpus Christi College MS F 291 offers some help in settling the question of how strongly were medieval dishes spiced. Based on its evidence, the degree of spiciness in different dishes seems to have varied

<sup>90</sup> These are BL MS Arundel 334 and the six recipes in the *Noble boke of cokery* which come from that source; at least one recipe in BL MS Sloane 1986 (edited as the *Liber cure cocorum* by Morris 1862) and in Norfolk, Holkham MS 674 (edited as the *Noble book off cookry* by Napier 1882); and three recipes in Bodleian MS e.Mus. 52 (edited by Hieatt 2008)

quite significantly, as for example the amount of spices in a recipe for pears in syrup would result in a much more spicier dish than we would be used to nowadays, while a recipe for capon braised with beef indicates very modest amounts of spices. The reason why Hieatt considers the latter to be the norm and the former the exception is not entirely clear, and the argument of Laurioux (2002: 19–20), namely that medieval dishes used more spices—both in variety and quantity—than we commonly do, seems entirely plausible and is also supported by records of the quantities of spices consumed by large households.<sup>91</sup>

Despite the concise and even elliptical nature of culinary writing described above, it seems unlikely that culinary texts were incomprehensible or useless to their intended audiences. Most recipes do include the ingredients for the dish, mention how they are to be prepared, mixed and cooked and even direct the reader to use particular kitchen utensils and containers. In addition, the recipes usually include quite specific information on the manner of garnishing and serving the finished dishes. It was simply a part of the professional skill of the cook to know the proper proportions of ingredients and the cooking time and heat for a dish of given type. The purpose of the recipe would then have been just to provide the cook with an understanding of the type of dish he was to prepare and to remind him of its details, not to provide technical instruction. Since it seems apparent that the terse style of culinary recipes is a conscious choice, it should not be viewed as merely a breach of Grice's maxim of quantity or quality, but rather a for implicit knowledge—knowledge that every medieval cook would possess as a result of his or her training as an apprentice, whether at home or in the great kitchen of a noble household.

### 8.3.4 Chronological evolution of medieval culinary recipes

Although 'medieval recipes' are here, as in elsewhere, treated as a single entity, neither the degree of information provided by them nor many of their other aspects remained constant throughout the medieval period. Both Carroll (1999) and Hieatt (1988) have observed that recipes translated from Anglo-Norman demonstrate a terser style than those originally composed in Middle English, and later 15<sup>th</sup>-century recipes tend to "spell out directions at considerably greater length" than earlier versions (Hieatt 1988: 16). According to Hammond (1993: 128), this chronological progression of the English recipe tradition from brief and often somewhat cryptic recipes towards more detailed ones is the result of an accumulative process, as consecutive users of the recipes added and expanded on earlier material. The end of 15<sup>th</sup> century seems to have marked the high point in the length and elaboration of recipes, as 16<sup>th</sup>-century authors copying medieval recipes usually represented them in shorter and simpler form (Hieatt 2012: 15). Of the known Middle English collections, Hieatt (2012: 15) has considered Corpus

<sup>91</sup> Although purely anecdotal, also my personal experiences in replicating medieval recipes would seem to support this view. Having replicated a number of recipes from the *PD* and other Middle English (and some continental) recipe collections using quantities of spices that allow their characteristic tastes to be appreciated but not to overpower the whole, the resulting taste combinations (with spices like pepper, cinnamon, cloves, ginger, saffron and cardamom being the most common and characteristic tastes) often resemble the cuisines of Middle East and many areas of Northern India, albeit without the burning element brought by chilli peppers.

Christi College MS F 291 to represent the extreme in this development, its recipes frequently being much longer than those in (supposedly) earlier collections; for example the recipe for *chauden for swan* in this manuscript is over 150 words long, the recipe for *caudle* 125 words long, and the recipe for *blancmanger* 192 words long, while most other collections have versions that are much shorter.<sup>92</sup>

Interestingly, this development also seems to be paralleled by a similar development in the dishes themselves. Generally speaking, the dishes described by 14<sup>th</sup>-century recipes tend to be somewhat plainer than their later versions, rarely calling for more than two or three spices for a particular dish. As recipes were passed down from one generation to the next, they tended to become not only more complicated, but also spicier and sweeter. Also the variety of spices seems to have increased with time, the earliest versions being limited to the basic pepper, ginger, cinnamon, cloves and nutmeg (Hieatt and Butler 1985: 10-2; Hammond 1993: 129). Especially the proliferation of sweeteners, such as honey, sugar and dried fruit, has been seen as a characteristically 15<sup>th</sup>-century feature (Hieatt 1988: 16, 2012: 14).<sup>93</sup> According to Strong (2002: 83), this increasing complexity and elaboration of dishes was an index “both of an increasing interest in food as an important aspect of court culture and of the arrival of a new class of nouveaux riches who regarded the arts of the table as one aspect of a new art of living” (see subsection 7.1.3).

This diachronic development does not seem to have been unique to England. The French Viandier de Taillevent provides an excellent example of this development due to its continued popularity and numerous surviving versions from different periods. According to Hieatt and Butler (1985: 11), Paul Aebischer—who has investigated the successive revisions of the Viandier recipes—has noted that each revision tends to be a little longer, spelling out procedures more carefully, occasionally suggesting additions or variations. Also Scully (1988) has noted that the earliest version of a cookbook existing in multiple copies is very likely to be the “shortest and most elementary of all eventual versions” (7).

Although written recipe collections served to codify and standardize recipes, their copying, transmission, adaptation and use also made possible a “much more rapid accumulation and wider diffusion of a record of successful culinary practice and experience” (Mennell 1986: 67), making them also agents of innovation and change. As was pointed out in subsection 3.1.2, this cumulative and evolutionary nature runs counter to the notion of corruption assumed by traditional textual theory, and means that changes undergone by recipes in the course of their transmission are just as likely to represent “legitimate and legitimating accommodations of a text to its readership over time” (Williams and Abbott 1999: 7) as anything that could be considered ‘corruption’. This means that the ‘best’ text in the traditional text-critical sense of being closest to the original is not necessarily the same as the gastronomically ‘best’ text (Scully 1988: 7). In fact, the whole concept of ‘best’ text, either from a culinary or text-critical viewpoint is not

<sup>92</sup> The different versions of *PD* also have recipes for these dishes, the ones for *chauden for swan* ranging from 105 to 116 words in length, for *caudle* from 54 to 59 words, and for *blancmanger* from 52 to 79 words.

<sup>93</sup> The recipes in Corpus Christi College MS F 291 seem to exhibit also these features of late 15<sup>th</sup>-century recipes “to an extreme degree”, including more sweeteners in its meat and fish recipes than any other collection known to Hieatt (2012).

very useful, since each version of the text reflects its particular cultural, situational and textual context and thus potentially quite different culinary tastes, pragmatic functions and intended audiences.

This does not, however, mean that ‘living texts’ such as recipes would be immune to corruption. The skills and understanding of medieval copyists and compilers—especially regarding culinary practice—were often far from perfect, and some recipes have merely become more confused rather than elaborated with time. However, if we are willing to give up the pursuit for ‘authorial intention’—as we should be—we can consider every user of the book—“every dilettante or practitioner, every gastronome chef or humble pastry baker” (Scully 1988: 8)—who made the slightest change in the text to be not only an *author* but to a certain extent also a *principal* who has provided his successors with what he considered to be a better text. As Scully (1988) recognized in his edition of the *Viandier de Taillevent*, this also has implications for the editing of these kinds of texts. Since no text can be objectively characterized as the ‘best’, and what we consider a single text may in fact have several ‘original’ versions based on the same culinary practices, there is no basis for a text-critical approach. This observation led both Scully and the present editor to see all of the surviving manuscript versions as presenting “texts which, while different, and imperfect in different ways, are individually valuable” (9) and to conclude that “there seems to be no good reason for paying less attention to any one” of them (9).

## 8.4 Producers and users of culinary recipes

Since the culturally defined purpose of a text—i.e. its genre—guides both its writing and the reading process, it is also intimately connected with the relationship between its authors and its audiences (Taavitsainen 2001: 89). If our intention is to study the relationships between the authors and audiences of medieval recipes, it is essential that we do not make any *a priori* assumptions about them based on what we know about the uses and users of modern recipes. In order to understand the nature of medieval recipes as documents and tools, it is important to review what is known about the kinds of *discourse communities* in which they most likely circulated and the types of people that produced, owned and used them. We should, however, also recognize that our established understanding of these discourse communities is largely based on interpretation of those selfsame texts and that we thus face a serious threat of circular reasoning. Although we must be careful of using earlier interpretations of the material to validate new ones, this section will nevertheless present some of the hypotheses and accepted theories about the nature of the authorship and readership of medieval culinary texts, if only as targets for validation or repudiation.

### 8.4.1 Recipe collections as evidence of historical practice

Although recipes can simplistically be considered as textual codifications of culinary practices, constituting a primary source of evidence for historical cookery, there has been some discussion as to what degree they reflect actual historical practice. The core issues in the discussion have been summarized by Grieco (1992:

29–30), who presents four separate and sequential questions that must be posed if we wish to use medieval recipe collections as historical sources:

- 1) What is the relationship between the cookbooks and actual culinary practice; were the dishes actually prepared or did the recipe collection have some other purpose?
- 2) If the recipes were actually prepared, can we say something about their target group? Who were the cookbooks meant for and used by?
- 3) In what context were the recipes prepared and eaten; are they for banquets, for everyday meals, for the sick, etc.?
- 4) What was the relative frequency with which different recipes were used? Were some more commonly prepared than others?

(After Grieco 1992: 29–30)

Based on his examination of the accounts for foodstuffs kept for the *Mensa della Signoria* of Florence in the light of contemporary recipe collections, Grieco (1992) himself answers the first question with the conclusion that “[i]t is quite certain that the recipes found in fourteenth and fifteenth-century cookbooks were not simply a literary exercise: they reflect a cuisine which was actually cooked and served” (37). Also Henisch (1978) has argued that medieval recipes reflect actual practice, based on the “practical hints and knowledgeable asides” (144) contained in them, but the most convincing argument for the genuinely practical nature of medieval culinary recipes can be considered to be their culinary success, which has been frequently attested to by several scholars who have actually gone to the trouble of applying culinary expertise to medieval recipes (see e.g. the reconstructions in Brears 2012).<sup>94</sup> The need for practical experimentation as tool in solving the editorial cruxes of medieval recipes has been frequently acknowledged for example by the late Constance B. Hieatt, and the present editor is also of the opinion that—adapting the words of Taylor (2004: 98)—if we insist calling medieval culinary recipes recipes, we should be prepared to cook them. The conclusion of the practical usefulness (and use) of medieval culinary recipes is now generally accepted by most food historians; for example Scully (1992: 23–24) considers surviving recipes to be the best single document of medieval culinary practices, not only because they can be taken to refer to real culinary practices, but also because each of them applies to a specific time and region, and because they have survived in relatively large numbers in contemporary manuscript copies.

However, as Scully (1997: 71) points out, we must be cautious in relying too heavily on recipe collections for evidence on what was *eaten* in the medieval period. First of all, recipes concentrate solely on food that was *prepared* in the medieval kitchen and contain little evidence about the consumption of either raw foods such as fresh fruit and vegetables, or foods that were produced commercially by specialized craftsmen or obtained ready-made. According to both Hieatt (1988: 19) and Scully (1992: 71), it is likely that for example fruits and vegetables appear so rarely in cookery books because they were so commonly served and easy to prepare, not because they were not eaten. For the same reason, especially common or basic preparations were often omitted, as they were assumed to have been learned by housewives from their mothers and by professional cooks from

<sup>94</sup> Also my own experiences in reconstructing selected recipes from the *Potage Dyvers* collection support this conclusion.

their masters (Paston-Williams 1993: 16).<sup>95</sup> Summarising this view, now generally accepted by culinary historians, Brears (2012) argues that “the recipes which were written down were largely those for which the memory had to be refreshed, as they were not in everyday use” (13): “Most basic processes, from plain boiling and roasting, frying bacon and eggs, making gruels and oatcakes, or even boiling boars’ heads, are therefore absent from early recipe collections, even though recorded in other manuscript sources.” (Brears 2012: 13.)

It must also be kept in mind that due to the socioeconomically uneven employment of written documentation, the recipes preserved in the manuscript collections are likely to reflect strictly aristocratic and upper-bourgeois practice (Scully 1992: 5). Even Braudel (1981), who sees the differences between the diets of the upper and lower classes as generally overemphasized (190-3),<sup>96</sup> concedes that “the diet of the peasants, the vast majority of the population, had nothing in common with the cookery books written for the rich” (187). Even in an aristocratic context, the more complicated dishes and showpieces—which often receive a disproportionate amount of attention in recipe collections—are considered to constitute more of an exception than the rule, meaning that recipe collections do not necessarily reflect the everyday diet of even the upper classes, but rather the kinds of dishes served on special occasions (White 2004b: 15). Instead of the spectacles that often receive the most attention in medieval recipe collections,<sup>97</sup> the everyday repertoire of the cook would have consisted of the simpler sops, stews, pies, tarts, flans, biscuits, roasts, sauces, jellies and puddings that also make up the bulk of medieval recipe collections (Scully 1992: 3, 255).

Lehmann (2003: 24) takes an even more marginal view of the recipes contained in medieval cookbooks, considering them to represent the more elaborate dishes that would be served only to the high table, the lesser guests at a feast being served more simple fare for which no recipes are provided. Lehmann sees this to be exemplified by the relative absence of recipes for beef (considered a gross food suitable for the lower classes) in the collections and the large number of recipes for different kinds of birds. It is naturally difficult to evaluate what constitutes a ‘relative absence’, but of the 380 unique recipes found in the collections of the *PD* family, 30 use beef as an ingredient (although in some cases merely in the form of broth used as a cooking base), individual collections containing anything from 11 (MS H279) to 21 recipes (MS As) using beef.<sup>98</sup> Despite being less frequent in recipes than pork, beef seems to have been consumed in large quantities in gentry

<sup>95</sup> For an example of this, see footnote 58 on page 276.

<sup>96</sup> He bases this view on the argument that in the 14<sup>th</sup> and 15<sup>th</sup> centuries the low population density across Europe meant that meat was plentifully available even to the lower classes and formed a significant part of their diet, and on his perception of even the food of the upper classes having been artlessly made.

<sup>97</sup> For example Heston Blumenthal’s recent television series focusing on the reconstruction of historical foods (*Heston’s Feasts*, 2009) focused its attention exclusively on such showpiece foods as fruit made out of meat, blackbirds baked into a pastry and the *cockatrice*, a fantasy animal consisting of half pig, half chicken (which the show mistakenly attributed to the Tudor period).

<sup>98</sup> Rather interestingly, this seems to make beef a more regular appearance in the recipes themselves than in the accompanying bills of fare, which contain no direct mentions of beef, unless the *grand chare* or *grosse chare* mentioned as a part of the first course of many feasts is interpreted to refer to a large roast of beef, although (Hieatt 1988: 18) reports the only known mention of beef in a menu known to her to occur in a menu served to commoners at the coronation feast of Richard III, found in BL MS Sloane 442.

households.<sup>99</sup> Beef does also appear in the list of provisions included at the head for the menu for a feast given to Richard II and John of Gaunt by the Bishop of Durham in 1387, found in MS H4016, which lists 14 salt oxen and two fresh oxen among the provisions required for the feast.<sup>100</sup> The most natural explanation for this discrepancy between the number of recipes using beef and its consumption is that the recipes—most of which are relatively simple pottages and other boiled dishes—are ones which were prepared in large quantities to be served to the multitude of servants and other personnel of lower status that made up the bulk of the people fed by a household.

### 8.4.2 Questions of authorship

As was observed in subsection 3.1.2 and has become obvious in this chapter, medieval recipes are best described as *living texts* in terms of their authorship and textual transmission; instead of deriving their textual authority from an individual author, they are rather based on the accumulated knowledge of a discourse community, consisting of “the residue of social occasions, in which unknown numbers of performers and redactors have participated [...] as equal participants in a continuing creative process” (Pearsall 1994: 125). This means, as was explained in subsections 2.3.1 and 3.1.2, that the prevailing humanist conception of authorship is simply inapplicable to Middle English recipes.

In the case of Middle English recipe collections—all of which are anonymous—the closest one comes to the use of an author as a source of authority, is the association of the recipes with prestigious *personæ* mentioned in the included bills of fare or in a colophon.<sup>101</sup> Even these are several degrees removed from any actual author, claiming merely that the recipes in question are based on the practice of someone associated with the household of the said *persona*. They say nothing of the ‘author’ of the recipes (if we indeed can imagine such a person), but do serve the function of associating the texts to a certain—rather prestigious—social milieu.

Since the production of texts nevertheless always involves one or more actual people, whose views, values and experiences they reflect to a lesser or greater degree, subsection 2.3.1 introduced a three-tiered division of the authorial role, proposed by Fairclough (1992: 78), into 1) the *animator*, or the actual writer of

<sup>99</sup> For example in the household of Dame Alice de Bryene—which Lehmann (2003) characterizes as “relatively modest” (24)—beef represented 48 per cent of the total meat consumption for the years 1418 and 1419, while pork represented 28 per cent, mutton 14 per cent, poultry 9 per cent and game of various kinds only one per cent (24). In the somewhat grander household of John de Vere, Earl of Oxford, beef represented 56 per cent of the total, pork 17 per cent, mutton 14 per cent, poultry 6 per cent and game 7 per cent (24).

<sup>100</sup> They would not, however, have accounted for a very large proportion of the meat, alongside 120 carcasses and heads of sheep, 12 boars, 14 calves, 140 pigs, 300 marrow bones, three tuns of salt venison, three does of fresh venison and several thousand poultry and fowls of various kinds.

<sup>101</sup> Two named attributions of a Middle English recipe collection are known. The best known is the headnote found in the BL MS Additional 5016 version of the *Forme of Cury*, which claims the collection to have been “compiled of the Maister Cokes of kyng Richard the Se[cu]nde kyng of [En]glond aftir the Conquest . the which was accounted þe best and ryallest vyaund[er] of alle cristen [k]ynges” and “compiled by assent and auysement of Maisters and [i.e. of] phisik and of philosophie þat dwellid in his court” (Hieatt and Butler 1985: 20). Another one, noted by Hieatt (1988) is a headnote found in the Yale University MS Beinecke 163 version of the *Ordinance of Pottage*, similar to the *FCon* but claiming that the collection was “compylett by maysters off fysyke and mayster cokys off Kyng Rycherd the fyrst” (13-4), which Hieatt finds “manifestly absurd” (14).

the physical text, 2) the *author*, or the person who puts the words together into a text, and 3) the *principal*, whose position is represented by the words. In the case of living texts like medieval culinary recipes, these positions are layered in the sense that they are potentially re-inhabited every time the text is copied. While any scribe copying any manuscript text is naturally an *animator*, every scribe that reformulates a recipe in his own words, adds clarifications or describes some procedure in more detail—as we know to have frequently happened—is also a new *author*. Furthermore, it could be argued that any cook or even a scribe who makes substantial changes to the dish described by the recipe or to the methods used to prepare the dish is essentially the *principal* of a new recipe (as a *work*) describing a new version of the dish. Thus each recipe represents not only the position of its original principal, but also of all those intermediate users who have modified the recipe—perhaps based on their own taste or practical convenience—before having it copied, and the language of every scribe who went beyond verbatim copying, whether in an attempt to clarify it or simply by substituting their own dialectal variants for those of the exemplar.

The general consensus among culinary historians seems to be that the ultimate originator of most medieval recipe collections was usually a professional cook, or at least someone working in a professional kitchen and with knowledge and practical experience of cooking (Wheaton 1983: 18; Scully 1992: 8). While the royal household is naturally the most prestigious context to associate a recipe collection with, it was by no means the only employer of professional cooks who could originate recipe collections, as most “[m]onasteries, manor houses, castles, and the houses of the wealthy bourgeoisie” would have employed at least one cook, and many cooks would have ran their own businesses (Weiss Adamson 2004: 57). Even if they did not originate in the royal court, medieval cookbooks “were compiled and copied by members of the educated elite, the clergy, nobility, and rich bourgeoisie”, which means that they are generally silent on the food of the lower classes, which made up the vast majority of medieval society (Weiss Adamson 2004: xvii–xviii).<sup>102</sup> However, while they may have originated in courtly kitchens, Lehmann (2003: 21) points out that medieval recipes could also be—and frequently were, as the case of the *Menagier de Paris* shows—adapted to suit the needs of different classes of users.

The fact that recipes were written down would at a first glance seem to suggest that cooks were literate, but as White (2004b: 15) points out, this cannot be assumed for the Middle Ages. Both professional and domestic cooks would have learned their trade in practice, either through apprenticeship or in their mother’s kitchen, and would not have needed written instructions for practising it, although as we saw in section 6.4, a literate professional like a cook would not have been all that unlikely in the late 15<sup>th</sup> century. However, as the threefold model of authorship used here implies, it is well possible for even an illiterate cook to have served as the principal—and perhaps even as the author, operating orally—of a recipe or an entire collection of them. Perhaps the most likely candidate for the original animator—and perhaps even the author, depending on whether he was

<sup>102</sup> This is not limited to English collections; many of the important European recipe collections—as well as the surviving Arabic ones—were compiled in a distinctly noble milieu: the *Viandier of Taillevent* at the French royal court, the *Buch von guter Spise* at the house of the bishop of Würzburg, and Chiquart’s *On Cookery* at the ducal court of Savoy (Scully 1992: 5).



writing from dictation or reporting the cook's descriptions of his activity in his own words—of most culinary recipe collections would have been one of a number of kitchen clerks—who would have been required to both read and write as part of their duties, or even the steward of the household, all of whom had daily interactions with the cook, who usually had to report his or her expenses to them on a daily basis (Weiss Adamson 2004: 58). One question which frequently arises in connection with historical recipes is their ultimate origin: “how far the writers of cookery books were themselves the originators of the recipes, and how far they had taken them from existing texts” (Wilson 2004: 19). The hypothetical chain of production described above is equally likely whether the recipes were original works or re-encodings of earlier recipes passed on to the cook from written or oral sources and perhaps modified by him before being again written down as new versions or even as new works. This process of appropriation and re-encoding is likely to have operated also across language borders, as for example Meredith (2004: 33) has seen it far more likely that the Middle English recipe collections were translations and adaptations from French ones than that they are original compositions. The fact that the textual histories of recipes most likely involved an oral component at one point or another makes them even more complicated to trace than in the case of more purely literate manuscript texts—living or otherwise—and makes the whole concept of an ‘original version’ very questionable.

Once the recipes had entered the written page, their dissemination was subject to the demands of potential readers and patrons just as any other text. The surviving version of a recipe thus usually has a long and involved history behind it even within the literary realm, not to speak of the oral tradition of unknown length underlying—and potentially even interrupting it. In addition to the complications resulting from the inherent instability of recipe texts, they are also made prone to transmission errors by the specialized nature of their content; as Hieatt has remarked on several occasions (1988: 12 and 1998: 133), it would not have been unusual for scribes copying culinary materials to misunderstand what they are copying and make mistakes, since “most scribes were clerics, not cooks” (Hieatt 1998a: 133) and “few of them could have known much about the art of courtly cookery” (12). According to Hieatt (1992: 19) it is evident that medieval English scribes often had no idea of the meaning of words or phrases that they were transcribing, and could produce results which were either incomprehensible—or worse yet—deceivingly understandable yet illogical and, upon closer examination, plain wrong.<sup>103</sup> Nevertheless, this does not mean that we should ignore Keiser's (1998b: 121) advice to pay attention to the scribes' response to any technical language contained in the text, as it can provide us information about their degree of familiarity with the topic matter and thus of the kind of discourse community in which the recipes circulated.

The nature of recipe collections as discourse colonies—i.e. compilations of individual textual units—also complicates the questions of authorship and origin, as

<sup>103</sup> According to Hieatt (1998a: 133), the same also applies to most lexicographers and editors of culinary texts, especially in the previous generations. Due to this, modern editions of culinary texts—especially ones produced before the last decades of the 20th century—should always be approached with a modicum of caution. The editors, just as well as the medieval copyists, may well have been misled in their interpretations, not realising that from a practical perspective, certain readings appear absurd and demand explanations that may not be obvious from the context.

their component parts could potentially have very different origins and textual traditions, being compiled from a variety of sources before being established as a collection. Based on the evidence of the surviving English recipe collections, even established collections would often get copied partially and combined with groups of recipes from other collections (Wilson 2004: 19).<sup>104</sup> This kind of recompilation could take place either as a conscious editorial activity by a ‘professional reader’<sup>105</sup> (Kerby-Fulton 2000: 103), aiming at the production of a new collection either for personal or commercial reasons, or by the accumulation of recipes from various sources by an amateur compiler.<sup>106</sup> The occurrence of this kind of ‘recompilation’ of recipe collections during their transmissional history is supported by the quite markedly varying content of the different versions of the *PD* collection, as well as by the existence of several collections which combine material from more than one ‘family’ of recipe collections (Hieatt 1992: 17-8, 2008: *passim*).<sup>107</sup>

This fragmentary transmission and recombinant nature of recipe collections means that the traditional critical goal of producing “a text as close as possible—textual witness(es) and good judgment permitting—to what came from the compiler’s pen and was the origin of the surviving textual tradition”, stated as the aim of editing vernacular scientific and practical treatises by (Keiser 1998c: 113), is simply invalid for such texts. This also means that while the origin of an individual recipe may very likely lie with the cook of some great household, the origin and ‘authorship’ of a specific *collection* need not be in any way related to an actual culinary practitioner. This multiple authorship and transmission of the medieval culinary recipe also means that it is often impossible to tell the stage of transmission at which certain of its features originate. The final text represents not only the understanding and level of competence of a single authorial scribe, but the cumulative (or subtractive) levels of competence of a string of individuals with various backgrounds.

<sup>104</sup> The tradition of copying recipes from earlier collections to build up a new one is not unique to manuscript texts, and did not die out immediately once cookery books began to be printed. For example *The Widowes Treasure* by John Partridge, published in 1585, claims to have been based on a household book compiled for private use by “a gentlewomen in the country” that Partridge had borrowed from a friend (Wilson 2004: 19), and as White (2004b: 16) notes, many recipes contained in printed recipe collections occur in several different collections, as authors freely borrowed them both from manuscripts and earlier printed collections.

<sup>105</sup> Kerby-Fulton (2000) defines a *professional reader* as someone “whose job it was to make decisions on behalf of the medieval reader about how the text should go down on the page—*conscious* decisions, that is, about editing, annotating, correcting, rubricating, or illustrating a text”—through “any number of interventionist or creative activities such as translating the original dialect, tinkering with the alliteration, suppressing or embellishing controversial content, imposing an unauthorized set of rubrics, or illustrating an episode contrary to what the text actually says—or said”, distinguishing them from mere scribes by restricting the term “to those who *knowingly* controlled what others would read” (Kerby-Fulton 2000: 103-4).

<sup>106</sup> Jones (2000: 33) has observed a this phenomenon in the context of vernacular medical recipes and other medical texts, which were also compiled into new composite ‘works’ by interested amateurs and non-practitioners.

<sup>107</sup> The Bodleian MS Rawlinson D 1222 version of *An Ordinance of Pottage* provides an example of a scribe ‘editing’ a recipe collection by reorganising and expanding it. According to (Hieatt 1988: 12), the scribe apparently began to copy the collection in its normal order (represented by several other manuscript versions with minor variations) but after copying a number of recipes decided to subdivide it into more detailed sections than his exemplar, and having done so, decided to supplement it with recipes from other sources, in some cases supplementing a recipe found in other MS versions by a new one borrowed from somewhere else.

### 8.4.3 Audience and function of medieval culinary recipes

Several food historians have brought up the closely intertwined questions of the *intended audience* and *function* of medieval recipe collections, but found very few certain answers. However, these questions are important for the analysis of the textual features of medieval recipes, because the answers we give to them—whether based on research or prejudice—necessarily affect the way we interpret these texts:

Very regularly our decisions about what is in the texts we choose to see is determined by what we do with the texts. Unfortunately, that all too often inhibits our ability to see clearly what the original makers and subsequent users of the text did with it. (Toon 1991: 92)

As was observed in chapter 7, in terms of cost and literacy, the potential readership of practical texts like recipe collections in the 15<sup>th</sup> century was perhaps wider than ever before. While England did not match other European countries in the production of *de luxe* manuscripts in the late Middle Ages, the lower end of the book market was very active, as “the ever-cheapening methods of manufacture and an increase in disposable income among the middle classes led the way towards an expansion of the manuscript book-trade”, creating a widening market which would eventually be filled by printed books (Meale 1989: 201). This growth of the potential market for manuscripts was “in part dependent upon and in part encouraged by” changes in the book trade, such as the replacement of parchment by paper, which resulted in a significant decrease in their price and a corresponding increase in their availability by the end of the 15th-century (Lyall 1989b: 26; Meale 1989: 217; Sponsler 2001: 4). In addition to the decrease in material costs for new books, also the flourishing trade of second-hand books made manuscript books available to a wider range of readers. (Bühler 1960: 33; Pearsall 1989: 6; Sponsler 2001: 4). However, it is important to remember that even at this time, books of any kind were still essentially a luxury commodity, owned in significant numbers only by the upper echelons of society:

With the exception of bibliophiliacs like Richard de Bury, who owned perhaps 1,500 volumes, eminent churchmen would rarely have owned more than a hundred books, and aristocratic owners, to judge from their inventories, many fewer. Sir John Paston, by chance one of the best-known collectors of the fifteenth century, owned only about twenty books or so. (Pearsall 1989: 7)

This means that the group of people who were likely to *own* books was probably much smaller than the group who would have been able to read them. However, recipe collections and other similar practical books would have been primarily household possessions instead of personal ones. The person owning a recipe collection would not have necessarily been the person making practical use of it: a collection of recipes might well have been commissioned by the head of a household for the use of his kitchen. Recipe collections can thus be considered to be a form of *communal literature* in the sense that they are not made use of primarily by individuals but by communal units such as households. This means that it may be more fruitful to conceive their intended audiences in terms of households—or kinds of households—rather than of individuals; while it may have been the head

of a household that made the decision to acquire a collection of recipes—either by commissioning one from a professional copyist, by copying recipes from one or more pre-existing collections he could borrow from friends, neighbours or relatives, or by having his own cook, clerk, or steward document the dishes already prepared in his kitchen—the selection of the specific recipes to be copied or written down could very well have been made by someone else with more culinary experience.

In terms of the kinds of households that were expected to make use of these recipe collections, we have little direct evidence. One rare indication of the social context of medieval recipe collections, pointed out by Carroll (1999: 29), is found in the closing formula of a recipe in BL MSS Additional 46919 and Cotton Julius D viii (edited by Hieatt and Butler (1985: 54)), which reads “& to þe lord vorþ bringen”, revealing an assumption that the dish in question (“*saumon gentil*”) is being served to a ‘lord’. However, this does not really tell us much—the term *lord* could equally refer to the head of a small merchant household than to a Peer of the Realm; either would be the ‘lord’ of their house.

In terms of food consumption it would seem that by the 15<sup>th</sup> century any household that could afford to acquire more than a few books could also afford the means to prepare at least some of the simpler dishes described in contemporary recipe collections, as food was in the late medieval period not only becoming “a central item of consumption by late medieval consumers, but an item increasingly available to more and more consumers, especially those living in towns and cities” Sponsler (2001: 14). Although many cookbooks are ascribed to the cooks of royal or ducal houses or contain internal evidence of an aristocratic audience (Carroll 1999: 29), it does not mean that they were necessarily written for princely audiences. For example the French *Menagier de Paris*, which was basically a humbled-down reworking of the more regal *Viandier of Taillevent*, is explicitly intended for a bourgeois audience and can be seen as an early sign of the social diffusion of the medieval *haute cuisine* down the social scale (Mennell 1986: 60):

The recipes of the urban-bourgeois *Ménagier de Paris*, for instance, resemble those of courtly cookbooks, suggesting an emulation of courtly eating by those in lower social strata, and some of the surviving English cooking books from the fifteenth century show a strikingly similar pattern. (Sponsler 2001: 15)

Although many of the recipes presented in medieval cookbooks would have been beyond the means of a more modest household, a fair amount of them would have been feasible for a special occasion even in a humbler cottage or as a part of the daily menu of a wealthier household (Redon, Sabban and Serventi 1998: 3–4), making the ownership of the book itself the more limiting factor. However, since they generally describe a cuisine that is significantly above subsistence level, recipe collections would have been useful mostly to people who could afford to have their instructions followed in their household, i.e. who had access to the required ingredients and labour and had the motivation to aspire to the outward manners of the upper classes. This would tend to limit their appeal to the nobility, gentry and prosperous middle classes who would have both the resources and the motivation for the kind of socially motivated consumption implied by the recipes.

While the fact that medieval recipe collections reflect an upper class culture—aspired to by the emerging middle class—is quite generally accepted by historians, a more thorny issue is presented by their actual use within the household. Mennell (1986: 65) has not only posed this question explicitly, but has also presented some of the more commonly considered options:

Were they [medieval cookbooks], for instance, written by practicing cooks for the use of fellow practitioners, or were they written as a record of high fashion by and for a literate élite who only vicariously commanded operations in the kitchen? (Mennell 1986: 65)

### Recipes for the professional cook

Many variants of these hypotheses about the intended function of medieval recipes have been presented by different scholars. Regardless of their view of the principal function of medieval recipe collections, most historians have accepted them as documentary evidence of the actual practices of medieval upper-class kitchens. Some scholars have also seen their most likely audience to have been professional cooks. For example Mennell (1986) sees the fact that most recipe collections—both printed and manuscript ones—are written in the vernacular as “a strong hint that they were written by practitioners for practitioners” (65), although by the 15<sup>th</sup> century this is not a very strong argument, as English would have been a perfectly valid—perhaps even the default—choice regardless of their intended audience. A more convincing argument for the professional nature of medieval recipes is their brief and terse nature, which would seem to indicate that the authors of these recipes “assumed that their ultimate readers might be journeymen or master cooks, too, who had no need to be taken by the hand like some first-day apprentice” (Scully 1992: 220). Considering the large amount of implicit knowledge they assume of their readers (see subsection 8.3.3 above), it seems obvious that unlike modern popular recipes published in cookbooks, points out, medieval recipes were most likely not intended to teach the novice how to cook (Lehmann 2003: 20; Carroll 2009: 58). This apparent failure of medieval recipes to live up to what is seen as the principal function of modern recipes has left many scholars questioning their purpose:

The recipes do not give precise instructions, nor any indication of quantity, so quite how they were used is a mystery. (Hammond 1993: 127-128)

If the receipts did not teach how to prepare the dishes, who used them? (Lehmann 2003: 21)

Another complication for the hypothesis that recipes were professional literature intended for the cook is constituted by the relatively general consensus among scholars that professional cooks did not need recipes to practice their craft, instead working “by memory, experience and training” (Hammond 1993: 127), and to suggest that a cook as “a trained craftsman and perhaps a recognized member of the cook’s guild, could prepare a meal only if a cookbook were propped open at his

elbow would have impugned his very dignity” (Scully 1992: 8).<sup>108</sup> This, together with the fact that most medieval recipes omit descriptions of basic processes but nevertheless contain significant detail on the spicing and colouring of the dishes prompted Hieatt and Butler (1985: 8) to suspect that at least the earliest English culinary recipes might be merely memoranda reminding the cook of the exact details of a particular version of a dish, rather than recipes in the sense of complete instructions for preparing it. Also Scully (1992: 25) has conjectured—based on the information contained in the majority of French and English recipes—that a professional cook might have used a recipe as a reminder of such details as the complete list of ingredients that should be included in a dish, the specific sequence of cooking steps, or the appropriate sauce for the dish. One useful way of viewing the medieval culinary recipe could thus be to see it from an intertextual—or even ‘interpractical’—viewpoint (Fairclough 1992: 85): not as an independent and self-sufficient text, but as a node of a textual network, whose meaning is not constituted not merely by the text itself, but also by all of the other texts brought into the process by the reader, i.e. other recipes and instructive texts, forms of cultural knowledge such as popular medical theories, and the whole of the cook’s professional and specialist knowledge.

### Recipes for the household administration

Another candidate for the users of medieval recipe collections is the group of various literate and educated servants employed in a medieval household: “reeves, factors, bureaucrats of all sorts” (Hanna 2011: 189). For example White (2004b: 15) has suggested that medieval recipe manuscripts could have been in the keeping of a literate kitchen clerk, who would have used them to advise the cook as necessary. Also Hammond (1993) presumes that recipes were not intended for the ordinary working cook, but rather for the household officials in charge of running the kitchen. His more specific hypothesis is that recipes would have been used to estimate the amount of ingredients that would be needed after a menu had been drafted (Hammond 1993: 128), but this seems unlikely since medieval recipes rarely give any indication of quantity, as Hammond himself also points out. However, the idea that written recipe collections could have been used by the officials running the kitchen is also accepted by Scully (1988), who sees it to be supported by the physical condition of many surviving manuscripts:

Useful cookbooks are used, by the household manager if not by the cook himself. They may even become dusted with powders, splattered with sauces, burnt or smudged with hot spoons used to hold them open. Probably more than any other type of book they are subject to the vagaries of an unprotected existence and the influences of an unfriendly environment. (Scully 1988: 13)

Although these kinds of working copies of medieval recipe books “do not seem to have survived very well” (Scully 1988: 13), they may very well have existed, as is demonstrated by the Valais manuscript (a parchment roll) of the

<sup>108</sup> The prologue of the *Du fait de cuisine* of Maistre Chiquart, chief cook to the Duke of Savoy in the early 15<sup>th</sup> century, contains an explicit acknowledgement of this, the cook author stating that he does not and has never owned a cookbook, working entirely from memory (Scully 1988: 13–4).

*Viandier*, which according to Scully (1988: 9) was “clearly a functioning kitchen cookbook—despite its several manifest errors—much rubbed and even splattered with grease”.<sup>109</sup> Scully also notes that working copies of culinary manuscripts would be likely to contain “interlineal and marginal notations”, “smudged, illegible words”, “intercalated favorite recipes”, and “expunctuations and cancellations” (Scully 1988: 14).<sup>110</sup> In contrast, some recipe manuscripts, like the three later copies of the *Viandier* edited by Scully (1988)—as well as some of the *PD* manuscripts—are clean and relatively neat, leading Scully (1988: 13-4) to the conclusion that they were “obviously library copies”. Another factor that can be seen as a strong indication that a recipe collection would have resided in the library instead of the kitchen is its inclusion in a miscellany manuscript containing other texts which would have been of no use in the kitchen, as in the case of MSS Ad and C.<sup>111</sup>

### Recipes for the master of the household

It has also been frequently suggested that recipe collections were not in fact used in the kitchen at all, but were produced solely for the benefit of the master or mistress of a household, serving to document the culinary standards of his or her household. An extreme version of the view of recipe collections as records of high fashion, summarized by Carroll (2009: 58), is that “the function of the medieval recipe was not to instruct at all, but was mere propaganda”, just like many modern coffee-table cookbooks presenting recipes by professional chefs, never intended to actually be prepared by the reader. A more moderate version of this view is represented by Scully (1992), who sees medieval recipes as a means of documenting “conspicuous consumption” (9) or of demonstrating the skills of the cook. According to Scully (1992: 25), it is entirely arguable that recipe collections were not compiled *for* the use of the professional cook, but only *by* him. The commissioner could well have been the master or mistress of a noble or a wealthy bourgeois household, the recipe collection functioning mainly as archival material (Scully 1992: 25):<sup>112</sup>

A recipe collection was compiled in manuscript form not for the cook in a noble or bourgeois household but for the master or mistress of that household. It served to document certain standards of an elite class. Occasionally revised with additions, deletions and modifica-

<sup>109</sup> Observing the manuscript, Scully also described it having the ink rubbed and the skin worn smooth particularly among the right margin between recipes 126 and 146, “as if fingers had frequently held the roll open at that point” (Scully 1988: 9). These kinds of potential signs of use in the kitchen are also found on several of the *PD* manuscripts, as indicated by the physical descriptions of the manuscripts in section 9.2 and the textual notes in the diplomatic transcripts (appendices B and D).

<sup>110</sup> In view of this, it is intriguing to note that two of the *PD* manuscripts—namely MSS As and C—also contain extensive corrections, additions and notations in another hand roughly contemporary to that of the original scribe’s. Some of these corrections (discussed more fully in the appropriate textual and explanatory notes of the transcription) seem to indicate a familiarity either with other manuscript copies of the same recipes or with actual cooking practices.

<sup>111</sup> Although the case of MS C is ambiguous in the sense that the recipe collection edited here and comprising the initial part of the current manuscript seems to have spent some time as a loose quire, showing considerable wear and soiling, which could be indicative of its being used or preserved outside of a library before it was bound in its current context.

<sup>112</sup> This view is also shared by Strong (2002), who claims that the surviving manuscripts “certainly did not belong to practising cooks but to those for whom they worked” (78).

tions, occasionally copied, with the approval of the master or mistress in order to please a flattering friend or relative, a manuscript collection of recipes reposed in the household library, not in its kitchen. With only the odd exception, these books are in a good, clean condition—a tribute, if not to the intrinsic value of the data they contain, at least to the cost of the material and labour that went into their making. Normally very few copies, between one and four on average, of each text survive today; this number undoubtedly reflects a very low number of contemporary copies made. (Scully 1992: 8)

Redon, Sabban and Serventi (1998: 4) see also the existence of recipe collections written in Latin to support this view, since they would have had a more limited accessibility than vernacular ones and were more likely to have been made for the master of the household than for the cook. Their hypothesis is that these recipe collections were used as a tool by the master of the household, enabling him to select suitable foods to every occasion and to communicate his wishes to his cook. This hypothesis, however, seems at odds with the format and structure of medieval recipes. If their purpose was merely the identification of dishes and communication between the master and his kitchen, would this function not be better served by a more descriptive approach to the dishes, describing their appearance, taste and other properties rather than providing sequential instructions for their preparations, which often give very little indication of what the end result will look or taste like? The highly procedural style of the recipes, together with the lack of description makes it unlikely that they were intended primarily for the purpose of identifying dishes; regardless of their intended audience, medieval recipe collections nevertheless seem to represent “first and foremost a technical literature” (Redon, Sabban and Serventi 1998: 1).

### **Recipes for the middle class**

One obvious solution for the apparent discrepancy between the highly technical and instructive character of the recipes and the professional cook’s lack of need for written instruction is, that they were indeed intended to provide practical instruction for cooking, only not for professional cooks. While the hypotheses presented above are all based on the use of recipe collections within large gentry or noble households employing a large number of specialized staff to cook not only for the head of the household and his immediate family, but also for a significant number of dependants and guests, recipe collections could also be intended primarily as means for smaller private households to emulate the culinary culture of larger and wealthier households. In a smaller bourgeois household at the lower end of the economic scale able to afford both books and selective cooking—e.g. that of a moderately wealthy urban merchant, burgher or tradesman, or of a country gentleman or a wealthy yeoman—where the lady of the house would most likely have been directly responsible for the running of the kitchen with the aid of a few servants, the functions of mistress, official and head cook would be vested on a single person, who would in fact be seen as an ideal target audience for recipes. While not possessing the extensive culinary experience and knowledge of the professional cook, a middle-class housewife would most likely have some experience of the basic procedures of cooking and of the preparation of basic dishes, requiring



instruction only on the finer points of spicing and presentation that are particular to noble cookery.

In addition to providing practical instruction, a recipe collection describing culinary preparations made in noble households would also serve as a source for ideas and for a sense of the general style of dishes that were fashionable and associated with high status, much as modern cookbooks by celebrity chefs or prestigious restaurants do.<sup>113</sup> Thus a recipe collections would have served a similar purpose to the other kinds of conduct literature (see subsection 7.1.3), namely of *commodifying* manners by “packaging behavior as a way of ensuring socio-economic mobility and personal happiness” (Sponsler 2001: 4–5), promising “lastynge blysse” and “self auauunce” to their readers, like the *Babees Book*, written around 1475 (Sponsler 2001: 5). Among the rising bourgeoisie, the ability to display not only the manners but the culinary habits of the upper classes would have provided a way of converting economic success into social prestige and aim for a higher status in society. However, the relatively small number of surviving copies of manuscript recipe collections means that if recipe collections were used for these kinds of purposes, this cannot have been a widespread phenomenon. But considering that the early printed recipe collections of the 16<sup>th</sup> century were explicitly marketed for middle-class housewives and the compilation of personal household books containing recipes also became common in the same period, it is not entirely unlikely that the roots of this phenomenon would go back to the late 15th century.

#### Codicological and contextual evidence of use

The type of book can tell us something of the wealth, and therefore status, of the original owner or owners, whether it is a professional production, or the work of an amateur scribe, perhaps intended for private use only.

(Jones 2000: 327)

The quality of writing and the physical attributes of the manuscript can also provide information about the intended audience and use of a book, and the reception and use of the collection can be studied via marginal notes, underlinings and other attention-drawing devices (Taavitsainen 2001: 91).<sup>114</sup> Based on the quality of handwriting in most of the extant Middle English culinary recipe collections, Hieatt (1998a: 134) is of the opinion that the volumes containing English recipe collections were not presentation copies intended to be pored over studiously like devotional or scholarly works, much less admired, but practical tools intended to be used as references or records of the personal interests of the owner. While none of the *PD* manuscripts could be characterized as ‘display texts’, there is still

<sup>113</sup>This would also explain the frequent presence of bills of fare describing the dishes served at the feasts of various notable personæ; they would provide a model to be emulated—even if on a much smaller scale—and allow the bourgeoisie a glimpse at the glamorous life of the upper classes and topics for conversation, much like the glossy magazines of today.

<sup>114</sup>In order to enable the systematic study of these types of devices, often impossible when working with traditional editions, the present edition records these kinds of paratextual annotations as faithfully as possible in a structured format (see chapters 10 and 11, especially subsections 11.4.2, 11.4.3 and 11.5.1).

quite significant variation in the quality and care of workmanship. While some of the more workaday versions of *PD*, like MSS C and D, are clearly intended for personal reference and may even have been copied by an amateur scribe for their own use, others, like MSS As and H279, with their regular and careful layout and neat professional hand could well be considered modest presentation copies suitable for a nobleman's library. Lehmann (2003) sees this physical diversity of medieval culinary manuscripts—ranging from utilitarian to borderline luxurious—to indicate not only a diverse readership with varying means, but also a variety of intended uses: “the luxurious versions might well be designed principally to record the splendour of court feasts, enhancing the prestige of the royal host”, while more utilitarian copies “were for use by an individual owner and his servants” (Lehmann 2003: 21).

While Taylor (1996: 51) has noted that books, “especially handsome ones, were treated with respect, especially by lay readers”, resulting in a paucity of marginal annotations in English texts, this degree of respect does not seem to have extended to the kinds of practical manuscripts containing culinary recipes. Like many manuscripts throughout the Middle Ages, also recipe collections frequently have annotations written in ink by their users:

These may include ‘pen trials’ (perhaps by the scribes using or copying the present manuscript), alphabets, prayers, requests for prayers of the readers, popular verses, drawings, scribbles, and doodles. Names of bishops, kings, and local religious are common (often adding clues to when and where a book was made or used). Among other things, they give the strong impression that the books were well and variously used. (Toon 1991: 84)

The extent of information that can be gleaned from these kinds of user annotation varies significantly; in most cases the annotator remains completely anonymous, and often even the function or meaning of an annotation remains unclear. In addition to these kinds of general annotations unrelated to the content of the manuscript, recipe collections also frequently contain annotations directly connected to the content of the collection, such as recipe numbers or other reference markers added by later readers, links between similar recipes or glosses of unfamiliar words. While pen trials and other non-related annotations merely indicate that the book happened to be at hand for a later writer, these kinds of content-specific annotations are a strong indication that the manuscript was still being actively read for its content at the time of the annotation, serving as an indication of its lifespan. For instance several of the *Potage Dyvers* manuscripts contain a large number of annotations both in 15<sup>th</sup>-century hands roughly contemporary with their writing and in various 16<sup>th</sup>-, 17<sup>th</sup>- and even 18<sup>th</sup>-century hands (see the subsection *Later additions* under the description of each manuscript version in section 9.2). In the best case—although not very often—these kinds of annotations can even provide hints about *how* the collection was being used or what the reader was particularly interested in.

In the case of recipe collections included in miscellany manuscripts, the textual context can also provide clues of their use. According to Lehmann (2003: 21), manuscripts containing exclusively culinary and other closely related material (like texts on table service etc.) might have been intended “for the staff of a

great household”, providing cooks with reminders of how to prepare the dishes or to inform household officers of the foodstuffs that the cook would need for the different dishes, while manuscripts where cookery is associated with medical matter could have been used by doctors or apothecaries for picking a suitable diet for a patient, and manuscripts where it occurred with other domestic lore could have served as practical guides to estate management.

While culinary recipe collections frequently occur with all these kinds of texts—and various others besides, as was shown in subsection 8.2.2—their association with medical material seems to be especially pervasive. Both Hieatt (1996) and Jones (2000) have observed that the inclusion of culinary recipes in an otherwise medical manuscript is quite typical, either as entire collections or individual recipes.<sup>115</sup> For some reason, Jones (2000) sees the inclusion of culinary recipes in an otherwise medical manuscript as suggesting that the manuscript “may have been made for the women of a household” (316), although it may equally well have been compiled for the use of a physician, considering that the prescription of suitable foods for a patient—either as a treatment for an illness or proactively—would have been an important part of the physician’s practice (see subsection 7.1.1) and a copy of a popular recipe collection would have been a useful reference work. While the association between English physicians and culinary recipes has not been studied, Weiss-Amer (1992: 78) has found evidence from Continental sources that physicians did occasionally play a part in the circulation of culinary manuscripts. As an example she mentions two culinary manuscripts edited by Carole Lambert for her doctoral dissertation (Lambert 1989: 27–34) that were copied for a French and a German physician.

As the preceding paragraphs have shown, the evidence presented for the different hypotheses on the intended audience of culinary recipes is relatively scant and far from conclusive. The main reasons for this are the scarcity of surviving copies that provide information about their original owner or early provenance and the lack of references to the use of recipe collections in records and other historical sources (apart from the brief remark in the Maistre Chiquart preface noted above). While various ordinances, household accounts and other instrumental texts, combined with archaeological evidence have provided us with relatively detailed information about the workings of a medieval household with regard to its food economy,<sup>116</sup> there seems to be remarkably little information about the role which recipe collections may have played in it. It may well be that future evidence will reveal to us the correct answer to the questions and hypotheses presented above and reveal the purpose for which medieval recipe collections were used and by whom, but judging from the variety of physical forms and textual contexts in which medieval culinary recipe collections survive, it seems equally likely that they were in fact used for a number of different purposes by a wider variety of people than is commonly believed.

<sup>115</sup> Hieatt (1996) mentions the BL MS Harley 5401 as an example of a manuscript containing a culinary recipe collection within “a fifteenth-century volume otherwise entirely devoted to medical matters” (54), while Jones (2000: 315) mentions BL MS Sloane 442, Wellcome MS 542, Oxford All Souls College MS 81, and BL MS Harley 2378 as examples of medical manuscripts that contain individual culinary recipes either as marginal additions or scattered among medical and other recipes.

<sup>116</sup> See Brears (2012) for a recent extensive survey on the topic.

## 8.5 Early printed recipe collections

While the tradition of the household manuscript miscellany was continued in the Early Modern period by the personal manuscript recipe books commonly kept by gentry families, containing culinary recipes, household hints of various kinds and medical remedies (Porter 1992: 2), recipe collections also started to be printed very early on. The fact that cookery books feature among the earliest printed books, and had been printed in most of the main languages of Western Europe by the middle of the 16<sup>th</sup> century, would seem to indicate an expansion in their potential readership down the social scale (Mennell 1986: 65):

With the advent of print, the cookbook, like other literary forms, opens up avenues of social mobility, for it is then possible to teach oneself the mores (the menus and table manners) of higher-status groups by studying books on cooking, household management or ‘etiquette’, just as one could ‘improve’ agricultural practices or religious beliefs by diffusing texts or tracts to others. (Goody 1982: 192)

The early printed cookery books of the 16<sup>th</sup> and 17<sup>th</sup> centuries were aimed mainly at women—housewives both poor and rich, ladies and gentlewomen—but “would also have been of use to professional cooks, in domestic service or commercial establishments” (White 2004a: 73). By the end of the 16<sup>th</sup> century, recipe books and books on household management were being printed in significant numbers which increased further over the next century. The popularity of these kinds of books is demonstrated by the repeated editions of some of these titles; for example Gervase Markham’s *The English House-wife* saw thirteen editions between 1615 and 1683 (White 2004a: 73), while the French *Viandier de Taillevent* was printed in at least fifteen editions between 1490 and 1604.<sup>117</sup>

When literacy increased and printed recipe books became affordable for a wider audience, the recipes in them came to include also ones reflecting “the simpler meals of all classes of society” (White 2007a: 9). Simultaneously, they also came to be used as a vehicle of social improvement; in the 16th century the printed cookbook, along with other manuals of domestic behaviour, helped the rising middle class to breach the old hierarchical organization of cuisine. As the titles of many such guides put it, they served to ‘reveal the secrets’ and ‘open the closets’ of rich aristocratic households,<sup>118</sup> and transformed the medieval social hierarchy and the opposition between high and low into one that was “more closely related to expenditure than to birth” (Goody 1982: 152). Furthermore, it seems that also the very nature of professional cookery and consequently also the whole discourse community of culinary professionals was undergoing change as a result of the proliferation of cookery manuals:

<sup>117</sup> According to Scully (1988), these printed editions “frequently contain such errors and mutilations of the earlier texts that their use by contemporary purchasers must have led to remarkable culinary adventures” (6).

<sup>118</sup> Examples of these kinds of household manuals from the 16<sup>th</sup> and 17<sup>th</sup> centuries include *The treasure of commodious conceits, & hidden secrets and may be called, the huswiues closet, of healthfull prouision* (1573), *The good husvives Jewell* (1596), the *Delights for ladies to adorne their persons, tables, closets and distillatories* (1603) and *The queens closet opened* (1663).

The written collection of recipes, and still more the printed book, had a number of interesting possible consequences within an increasingly literate circle of professional cooks. First of all, they broke the absolute dependence of the transmission of culinary knowledge and skill on apprenticeship and direct personal relationship, and make possible a wider transmission of knowledge than any oral tradition of word and gesture. Secondly, writing down a recipe tends to enhance its prescriptive character; the imperative tone of early recipes is very striking. (Mennell 1986: 67)

## 8.6 Conclusion

Although recipes have been seen as an unusually stable genre and text type,<sup>119</sup> the comparison of the archetypal medieval recipe to an archetypal modern one makes it clear that there have been significant changes in the text type of recipes. Based on the above discussion and the contextual framework presented in section 2.2, it would seem that these changes are partially due to changes in the *context of culture*, which has changed the status of cooking as a skill, and partially due to changes in the situational context—partly caused by the cultural changes—in which recipes are used. As the cookery in the culinary sense, referring to the preparation of dishes with more than nutritional value, spread from the master cooks of great households to the cooks of more modest households and to housewives doing their own cooking, and finally in the recent past to people taking it up as a recreational pastime, the cultural function of recipes also changed from ‘providing information necessary for the preparation of a dish’ to ‘providing the information necessary for the preparation of a dish’, turning them from *aides-mémoires* for experienced professionals to tutorials for the beginner. Similarly, the context in which recipes were intended to be used widened from a large-scale professional context with established practices and experienced personnel to people learning to cook for the first time in a small-scale domestic setting with little information about the procedures involved. The elucidation of the chronology of these changes and the historical processes underlying them is an interesting research topic, and will require not only continuing historical research into the cultural and situational circumstances of cooking in different periods, but also an extensive diachronic corpus of recipe texts from the Middle Ages to the present day, which will allow us to track the chronology of the changes that have taken place in the text type of recipes.

In terms of recipes as a text-type, one relatively obvious topic that has not—rather surprisingly—been sufficiently studied is their lexis. As Carroll (1999: 32) points out, for example Görlach’s 1992 analysis of the recipe genre does not address the issue of text-type specific lexis, even though there are Middle English words—like *parboilen*—that are attested only in recipes and others—like *sethen* and *blaunchen*—that are used in a more specific or technical sense in recipes than elsewhere. Unfortunately, Carroll’s call for a “more thorough study of this aspect of vocabulary [...] in order to determine the extent to which lexical choice in reci-

<sup>119</sup> For example Görlach (1992: 756) has argued that cookery recipes have “seen less development than other types” in the course of their history.

pes is restricted by text-type (recipe) rather than topic (cooking)” (Carroll 1999: 32) has so far gone largely unheeded.<sup>120</sup> The purpose of the present edition is to provide, first of all, a resource for the detailed study of the language of Middle English culinary recipes in its manuscript context, and secondly, to provide a general model for the production of further such resources, which would in time form the kind of long diachronic corpus required for the study of the development of the recipe genre.<sup>121</sup>

---

<sup>120</sup> Exceptions to this include a brief article by Meredith (2004), which focuses mostly on the etymology and language of origin of the lexis of a selection of recipes from the *FC* and the *DC* collections (edited in Hieatt and Butler 1985), and the present author’s article on the diachronic development of cutting verbs used in culinary recipes from the 14<sup>th</sup> to the 19<sup>th</sup> centuries (Marttila 2009). In addition to these published articles, an unpublished conference presentation by the present author, titled “‘Gode’, ‘faire’ and ‘fresshe’ food: adjectives and their collocation patterns in Middle English culinary recipes of the Potage Dyvers family” and presented at the 16th International Conference on English Historical Linguistics in Pécs in 2010, analysed the different types of adjectives used in culinary recipes and their various functions.

<sup>121</sup> While a thorough study comparing culinary recipes to recipes representing other fields, which would answer the question posed by Carroll is beyond the scope of this thesis, the present edition, together with existing corpora and editions of medieval and early modern medical and other types of recipes (such as the *Middle English Medical Texts* and *Early Modern English Medical Texts* corpora) will allow this kind of a systematic study to be conducted using modern corpus linguistic methods.

## **Part III**

# **The edition**





## Chapter 9

# The *Potage Dyvers* family of recipe collections

The title *Potage Dyvers* was first associated with the group of fifteenth-century culinary recipe collections edited here by Hieatt (1992) in her account of the surviving medieval English culinary recipe collections, where she identified seven ‘families’ of recipe collections that survive in more than one copy (see subsection 8.1.1). As White (2007b: 21) points out, this title is not in fact used for the whole of the collection in any of the six versions, but is actually the title given to the first—and principal—subsection of the collection found in British Library MS Harley 279. In her article, Hieatt used it as a referential label for the ‘family’ of collections consisting of the more or less closely related Bodleian Library MSS Ashmole 1439 (As) and Douce 55 (D), and British Library MSS Harley 279 (H279) and Harley 4016 (H4016), which were originally edited by Austin (1888). Although Hieatt herself did not persist in the use of this title in her later work (e.g. 1996 and 2004), preferring more descriptive labels like “the manuscripts printed or collated in Thomas Austin’s *Two Fifteenth-Century Cookery Books*” (Hieatt 1996) or “the group edited by Thomas Austin” (Hieatt 2004) instead, it has here been adopted—frequently abbreviated as *PD*—as a convenient way of collectively referring to the six related recipe collections included in this edition.

### 9.1 Identity of the *Potage Dyvers* family

While the title of the *Potage Dyvers* family and its identity as a ‘family’ were established by Hieatt (1992), the existence of close parallels between some of the manuscripts was noted already by Austin (1888), who edited the four manuscripts mentioned above as two separate collections, collating together MSS As and H279 and MSS D and H4016. In her 2004 article, Hieatt identified a fifth member of the same family, found in British Library MS Additional 5467 (Ad), a miscellany manuscript containing a variety of practical and historical material, and pointed out that while MSS As and H279 were correctly identified by Austin as being related closely enough to be considered as versions of the same collection, MSS D and

H4016 are in fact quite different, especially in terms of their organization. Based on her examination of MS Ad she concluded it to be very similar to MS D, the two MSS constituting the titular “third fifteenth-century cookery book” in addition to the two Harley manuscripts edited by Austin.

The sixth manuscript collection that clearly belongs to the *Potage Dyvers* family as defined by Hieatt (1992) was discovered by the present author while editing the recipe collection contained in part A of MS Cosin V.iii.11 from the Durham University Library (MS C) for a Master’s thesis (Marttila 2006), when the collection turned out to be a close—albeit not identical—parallel to MS H4016. The identity and extent of the *Potage Dyvers* family as understood here is thus based exclusively on earlier research, and from a philological point of view one of the purposes of this thesis is to evaluate the sensibility of considering these collections as members of the same ‘family’, as well as to establish their precise familial relationships in terms of their textual content and structure. The basis of the family’s identity in earlier research also means that the membership of the family cannot be conclusively be limited to these six manuscripts, as MSS that have hitherto not been properly analysed could well prove to be sufficiently similar to the MSS edited here to be considered members of the *Potage Dyvers* family.

As will be shown below in section 9.3, there exist several known manuscripts which have not been considered as members of the family, but nevertheless share significant amounts of recipes with it and could well be argued to belong to it. This indeterminacy in the identity of recipe collections—which are often catalogued only by the incipit of their first recipe (Mooney 1998: 124) and thus hard to identify from library catalogues alone—means that they benefit greatly from the *extensibility* of digital editions: even if the difficult task of locating all the relevant manuscript versions—generally held to be the first challenge of a manuscript editor (Keiser 1998c: 110)—is left incomplete, transcriptions of newly discovered versions can be integrated into the edition even after its completion.

Additionally, the nature of the *Potage Dyvers* text as a *discourse colony* has implications for the analysis of its transmission history and thus its identity. Based both on Hoey’s (1986 and 1986) observations on other types of discourse colonies and on the variation in the selection and ordering of recipes in the six versions, as well the occurrence of some of the recipes also in other, less closely related collections, the transmission histories of its component parts are likely to differ significantly. This also means that the identity of the *Potage Dyvers* ‘family’ of recipe collections is very much a *post-hoc* construction of textual scholarship, and it is difficult to evaluate to which degree this family of collections was perceived as an established whole when it was copied in the 15<sup>th</sup> century. The lack of a consistent title would seem to indicate that it was most likely not seen as a collection with a specific named identity distinct from other recipe collections circulating at the time, but its identity may have been defined simply through its manuscript context—the fact that the recipes had been collected together by someone—strengthened by the table of contents included in most of the versions. Its reader would not have seen it as a member of *the* family of collections now known as *Potage Dyvers*, but simply as *a* collection of recipes, its identity being established gradually as it was repeatedly copied in a more or less similar form.

## 9.2 The manuscript versions

As has been mentioned, the recipe collection commonly known as *Potage Dyvers* (after Hieatt 1992: 21) is known to survive in six manuscript versions, held in three different libraries. Each of these versions is unique in the sense of not being a mere scribal copy of another surviving member of the group. The similarities and differences between them vary, both on the level of recipes included, their textual organization, and their linguistic and visual presentation. The different versions also occur in varying manuscript contexts: two of the six versions occur as a part of a larger miscellany or a household book containing a variety of mostly medical and otherwise practical—but also religious and historical—material, while four survive as independent books.<sup>1</sup> Three of the versions are written on parchment, while the other three are set down on paper. Although Lyall (1989b: 12) has concluded that the use of paper was very rare in the beginning of the 15<sup>th</sup> century, paper accounting for about 20 per cent of all surviving MSS from the middle of the 15<sup>th</sup> century and about 50 per cent or more in the final decades of the century, there seems to be no distinction between paper and parchment in the MSS of the *PD* family, both having been dated equally to the beginning, middle and end of the century. This might be partially explained by the fact that even all of the three parchment manuscripts are written on rough parchment of low quality with various kinds of imperfections, which would have been relatively cheap in a sheep-rearing country such as England (Caie 2008: 16) and quite typical to the lower end of the book market (Lyall 1989b: 14). The lower end books would have also been a natural context for the early appropriation of paper, the likely option for recipe collections and other practical manuscripts “at least from about 1450 or so” (13), or even from the late 14<sup>th</sup> century in the case of informal documents such as household accounts, correspondence and “memoranda of all kinds” (Hector 1966: 17).

The scribal hands range from an uneven cursive anglicana-secretary hybrid to a careful *textualis semiquadrata*, and the decoration of the manuscripts varies from nonexistent to the liberal use of rubricated initials and line-fillers, most likely intended primarily as text-organising devices and finding aids. While the hand and material used for the manuscripts do correlate to the extent that the two versions written in a *textualis* hand are written on parchment, the third most formal hand—writing in a formal hybrid secretary script—occurs in a paper manuscript (H4016) and the third parchment MS (D) is written in a relatively cursive hand.<sup>4</sup> The decoration follows a similar pattern, as two of the paper manuscripts are noticeably less

<sup>1</sup> However, since none of the MSS retain their medieval bindings, this does not mean that also the currently individual MSS could not at some point have existed as a part of a miscellany, as for example Thomas Rawlinson (1681–1725), a barrister and a noted bibliophile, had the habit of dismembering manuscripts, breaking miscellanies into individual texts (Allen 1987: 7), and all of the individually bound *PD* versions apart from D are short enough to have once been a part of a miscellany.

<sup>2</sup> The recipe collection, including the accompanying tables of contents and bills of fare, occupies 44 folia of this miscellany manuscript.

<sup>3</sup> The recipe collection, including the accompanying tables of contents and bills of fare, occupies 21 folia of this miscellany manuscript.

<sup>4</sup> This MS is also distinguished by its very small size, which may mean that it was made of smaller and cheaper leftover sheets.

	Harley 4016	Harley 279	Add. 5467	Ashmole 1439	Douce 55	Cosin V.iii.11
Dating	mid 15 <sup>th</sup> c.	early 15 <sup>th</sup> c.	late 15 <sup>th</sup> c.	late 15 <sup>th</sup> c.	mid 15 <sup>th</sup> c.	early 15 <sup>th</sup> c.
Material	paper	parchment	paper	parchment	parchment	paper
Folio size (mm)	280 × 195 (Royal quarto)	220 × 150	210 × 145 (Chancery quarto)	210 × 140	135 × 95	220 × 150 (Chancery quarto)
No. of folia	29	50	224 <sup>2</sup>	55	80	92 <sup>3</sup>
Text block size (mm)	205 × 140	160 × 105	155 × 95	150 × 100	90 × 60	180 × 115
Lines	30	29	~30	~26	~20	~36
Scribal hand	formal hybrid secretary	<i>textualis</i> <i>semi-</i> <i>quadrata</i>	semi- cursive secretary- anglicana hybrid	compact <i>textualis</i> <i>semi-</i> <i>quadrata</i>	compact cursive secretary- anglicana hybrid	semi- cursive anglicana- secretary hybrid
Minim height	c. 3 mm	c. 2.5 mm	1-2 mm	2-3 mm	c. 1.5 mm	c. 2 mm
Rubrication	Paraphs, frames, line fillers and highlights.	Titles, initials, frames, line fillers and highlights.	None.	Titles, line fillers and highlights.	Titles, recipe numbers, large initials and paraphs.	Paraphs and highlight- ing.

**Table 9.1:** Overview of the physical features of the manuscripts containing a version of the Potage Dyvers collection.

decorated than the parchment ones, MS H4016 again being the exception.<sup>5</sup> These differences in the physical properties of the six MSS reflect the fact that different manuscript versions of a text are never on an equal footing, but rather reflect “substantially different social sitings”, being copied from different exemplars and for different purposes and audiences (Hanna 1992: 122).

All of the manuscript witnesses for the *Potage Dyvers* collection have been dated to the 15<sup>th</sup> century. In their concordance of English recipes, Hieatt and Nutter (2006) provide approximate dates for all of the manuscripts listed, including five of the six *Potage Dyvers* versions edited here, apparently mainly on the basis of Austin (1888). The dates given by are ca. 1430 for As, ca. 1435 for H279, ca. 1445 for H4016, ca. 1450 for D and Ad. Based on—unfortunately relatively scant—internal evidence found in the manuscripts and the estimates of various cataloguers, discussed below in the subsections for the individual manuscript versions, these datings have been revised somewhat, the dates proposed here being summarized in *Table 9.1*. Because of the scarcity and inconclusive nature of the evidence, the level of detail in dating the manuscripts has been restricted to characterising the manuscripts as being most likely from the *early* (1400-1440), *mid* (1440-1460) or *late* (1460-1500) 15<sup>th</sup> century, although the emphasis here must be placed firmly on “most likely”. Furthermore, as Hieatt (1996: 56) points out, the

<sup>5</sup> The dating of MS H4016 to the middle of the century would make it an unusually early example of a relatively handsome and formally executed manuscript written on paper.

date of the surviving manuscripts does not always tell much about the antiquity of the recipes contained by them, since material dating from earlier centuries was often incorporated in fifteenth-century volumes, and for example in the case of *Potage Dyvers*, the long transmission history suggested by the extensive variation in the extent and ordering of recipes within the six surviving versions means that the recipes themselves—or at least their precursor forms—can well be much older than the manuscripts.

In order to establish the various manuscript contexts in which the recipe collection has survived in, the manuscripts are described in terms of their physical makeup, visual layout and decoration, intellectual content, scribal hands and their provenance, as far as it is known. The information about the manuscripts contained in this section is also included in structured format as a part of the metadata header (see section 11.1) accompanying the TEI XML base data files included in this edition (see appendix A).<sup>6</sup> *Table 9.1* provides a quick overview of the physical properties of each manuscript. The complex structural and linguistic relationships between these six manuscript versions of the *Potage Dyvers* collection as composite entities—or discourse colonies—are analysed in chapters 12 and 13.

## 9.2.1 London, British Library MS Harley 4016

### Physical description

The manuscript is written on cream coloured, finely textured yet sturdy paper stock and is made up of 29 medieval paper folia, surrounded by 6 modern paper leaves. The laid lines run vertically on the page, indicating a quarto format, and the chain lines cannot be discerned. There are no watermarks discernible by the naked eye. The folia measure 282 mm in height and 195 mm in width, the typical size for a slightly trimmed Royal quarto (Gumbert 2000: 81-2), and the thickness of this rather slender manuscript, excluding the covers, is merely 6–8 mm, with the covers adding another 6 mm. The paper is sturdy and opaque enough that the writing on the opposite side of the page is only very slightly visible through it, although the red pigment used for decoration shows more evidence of bleed-through. The manuscript is generally in good condition, but there is some moisture damage evident on the first and second folia of the manuscript, f. 1 being quite clearly darkened with some distinct waterlines and f. 2 having some darkening around the spine, with no distinct waterlines.

### Collation of the manuscript

In the current binding, the original manuscript quires are both preceded and followed by two modern paper flyleaves and a single thinner, possibly earlier modern folio separating the flyleaves from the original folia. The collation of the manuscript, excluding the above-mentioned modern leaves, is as follows:

<sup>6</sup> The descriptions of the six manuscripts included here (and in appendix F) are based on the contents of the <msDesc> element included in the *TEI Header* of each of the source TEI XML documents. The data contained in the header was first automatically converted to appropriately titled sections of prose paragraphs and lists, and subsequently copy edited manually to remove redundancy resulting from the structured format of the source presentation and to make it more appropriate for prose presentation.

- 1<sup>8</sup> (ff. 1-8);
- 2<sup>8</sup> (ff. 9-16);
- 3<sup>8</sup> (ff. 17-24);
- 4<sup>6</sup>, wants 6, the final modern folia being glued to the stub left by it (ff. 25-29);

**Signatures and catchwords** There would seem to be two sets of signatures at the extreme lower right edges of the recto sides throughout the manuscript, partially on top of each other: one in the same brown ink that has been used for ruling the pages (hand 1), and another later one, written partially on top of it in red pigment (hand 9). On some pages, these signatures have been partially lost, most likely due to trimming of the pages. The signatures in brown ink consist of the letter *a* on ff. 1 and 4, the letter *b* on ff. 9 and 12, the letter *c* on ff. 17-20 and 22, and the letter *d* on ff. 26-28 (with traces of a signature at the very edge of f. 25). In the red pigment, ff. 1, 3 and 4 are signed *ai*, *a.iiij* and *a.iiij*. (the signature on f. 2 being supposedly lost due to trimming); ff. 9-12 are signed *bj*, *bij.*, *b* and *b.iiij* (the number on f. 11 lost due to trimming); ff. 17-20 are signed from *cj* to *c.iiij*; and ff. 25-28 are signed from *dj* to *d.iiij*, which would point to the scribe inserting these later signatures not having realized that quire 4 only ever consisted of three bifolia, with the sewing between ff. 27 and 28. There are also catchwords, written in the same hand as the text itself on the right side of the bottom margin, on f. 8v ("browes"), f. 16v ("brothe"), and f. 24v ("and salt"). All of the catchwords are surrounded by a frame drawn in red pigment.

**Foliation** The first 28 original folia of the manuscript have been foliated in an 18<sup>th</sup>-century hand writing in black ink (hand 4), possibly by Edward Harley. The folio numbers are located in the top right hand corner of each recto page and consist of an arabic numeral surrounded by parentheses. In addition to the folio numbers, ff. 1r, 9r, 17r and 25r (i.e. the first pages of each quire) have another number in the same hand, running from 37 to 40, placed both above or below the folio number, separated from it by a horizontal line, and at the center of the bottom margin. The frame of reference for this quire numbering is unclear, but it might indicate the location of these quires within a particular set of manuscript materials acquired by Harley as a single collection or purchase.

### Manuscript layout

The layout of the original manuscript folia is very uniform, consisting of a single column of text outlined in brown ink (hand 1) by vertical and horizontal lines extending all the way to the edges of the page and lineated by 28 horizontal lines which frequently extend past the vertical lines but not to the edge of the page. The text block measures 200-205 mm in height and 135-140 mm in width. Ff. 1r-2v also have a third vertical line down the middle of the text block, dividing the page into two columns, although the text has been written in two columns only up to the middle of f. 2v. The remaining space is divided between the margins, the top margin measuring c. 25 mm (to the top line), the bottom margin measuring 55-60 mm and the inner and outer margins measuring 20-30 mm (including a gutter of c. 5 mm) and 28-30 mm, respectively.

The ruling is very consistent, but there is no evidence of pricking in either the inner or outer edge of the leaves. The absence of pricking is corroborated by the fact that the rulings on the recto and verso of a single folio very rarely line up, indicating that the ruling was done separately for each page (or side of a bifolium), possibly using a template. The ruling is quite strictly observed by the scribe, although ends of lines do occasionally extend past the right-hand side of the ruled block, and recipe-initial paraphs are placed in the left-hand margin. Incomplete lines are filled to the edge of the text block with a pattern of slanting strokes in red pigment. The first line of text is written above the top line, resulting in 30 lines of text on each page. The height of each line is c. 7 mm, and the text has been written with the baseline on the ruled line, descenders extending below it. The minim height of the scribal hand is very consistently c. 3 mm.

In terms of marking textual organization, each recipe is separated from the preceding one by an empty line, on which the recipe title is written, in or near the center of the page. The table of contents, laid out in two columns, has no recipe numbers, but the empty space left after the recipe title is filled with a line filler consisting of curved strokes in red pigment.

### Current binding

Judging from the information on the binding, the manuscript was most likely bound into its current binding sometime after its sale to the British Museum in 1753. The manuscript is bound in brown cloth, with the spine and c. 40 mm of the spine edge and the front corners reinforced in brown leather, the covers measuring 290 mm in height and 205 mm in width. The coat of arms of the Harley family—"Or, a bend cottised Sable"—supported by two angels, surmounted by the coronet of an Earl and accompanied by the motto "VIRTUTE ET FIDE" decorate the front and back covers. On the spine, there are six cords dividing the spine into seven compartments into which the following text has been printed in gold (except for the shelfmarks at the ends of the spine, which are applied on white paper stickers): "61 - A BOKE OF KOKERY. - BRIT. MUS. - HARLEY MS 4016. - - - B.12".

### Summary contents of the manuscript<sup>7</sup>

- 1) Two menus (ff. 1r-2r)
  - a) A menu of three courses served by the bishop of Durham to Richard II, with a list of provisions (ff. 1r-1v)
  - b) A menu of three courses served at the consecration of the Archbishop of Canterbury (ff. 1v-2r)
- 2) A collection of 182 culinary recipes (ff. 2r-28r)

### Hands used in the manuscript

The manuscript is written entirely in a single hand and has very little annotation, apart from a foliation and some notes by cataloguers or librarians.

<sup>7</sup> A more detailed description of the contents of this MS is contained in appendix F and in the TEI header of the base data file for the MS included in appendix A.

**Scribal hand** The original scribal hand is a moderately formal and neat secretary hybrid (*littera cursiva secretaria media*) slanted slightly to the right, with some anglicana features. The ink used is dark and crisp greyish black in colour, occasionally fading to lighter grey where the ink has run thinner. The headings of the recipes are often written slightly more carefully and formally. The height of the letters is very consistent, the minims being about 3 mm in height. Interlineal additions made into the text would seem to be in the same ink and hand, although they are too small and delicate to be certain. The typically secretary features exhibited by this hand include the *d* with the straight diagonal ascender (although the anglicana *d* with the looped ascender also appears occasionally), the *g* with the horizontal top stroke and open curved descender (although the 8-shaped anglicana variant also occurs on some headings), the kidney-shaped round *s*, which is the default choice in word-final positions, and the ‘lobed’ secretary *e* with a rather pointed top. The only conspicuous and consistent anglicana feature is the exclusive use of the two-compartment *a*, in addition to which some letterforms display the mixed quality of the hand. These include the *r*, which occurs seemingly indiscriminately in both the long anglicana and the short secretary forms, and the *w*, which is of a consistent shape which lies somewhere between the elaborate anglicana form and the simple secretary one, consisting of two straight diagonal strokes followed by a double loop (instead of the single one more typical of secretary). The script seems to distinguish between the letters *z* and *3* by having a small curl at the bottom of the former. The letter *i* has occasionally been indicated by a short diagonal stroke above it or slightly to the right, especially next to letters consisting of minims.

**Other hands** In addition to the scribal hand, the following hands can be identified in the manuscript:

- 1) A 15<sup>th</sup>-century hand in brown ink that has prepared the pages for writing by ruling the writing block and writing the original folio signatures to the bottom corners of the pages, and apparently also added some markings into the margins of some pages.
- 2) A 15<sup>th</sup>-century hand in grayish brown ink added some superscript *a* symbols into the margins next to some recipes. The ink used by this hand is very similar to that used by hand 1.
- 3) A 15<sup>th</sup>-century hand in red pigment that has added decoration, line fillers and highlighting to the manuscript. Could belong to the original scribe, but since this hand has not written any text, it is impossible to determine the identity of these two hands.
- 4) An 18<sup>th</sup>-century hand in faint black ink, possibly belonging to either Edward Harley or his acting librarian at the time, David Casley, that has added the acquisition note on f. 1r, the folio and quarto numbers onto the recto side of each folio and the closing note on f. 28r.
- 5) A late 19<sup>th</sup>-century hand in pencil that has added the note on f. 29r and possibly also the formula “2/III e” on the second line of the second flyleaf.
- 6) A modern italic hand in reddish brown ink written with a wide nib that has added the shelfmark “123.B.14” to the first line on the second flyleaf. The hand could possibly belong either to Edward Harley himself or his acting librarian at the time, David Casley, the deputy-keeper of the Cotton library.
- 7) A modern italic hand in purplish black ink written with a narrow nib that has added the catalogue number “4016.” and a horizontal line above and below it to the second line on the second flyleaf. This hand could belong to David Casley, who catalogued the manuscript.



## Decoration

The predominant form of decoration found in the manuscript is the use of red pigment, both for highlighting text written in ink and inserting various text-structuring elements. The most prominent of these are the paraps indicating the beginning of a new recipe, inserted in the left margin. They are just under two lines in height and located next to the first and second lines of each recipe. The other form of text-structuring decoration are the line fillers, consisting of a sequence of wavy strokes slanting to the left and straight oblique strokes slanting to the right, used to fill out any incomplete final lines of recipes and to signal the end of the recipe.

Red pigment is also used to highlight not only the titles of the recipes, which are more or less centred in the writing block and ‘cradled’ by a red box open at the top, but also individual initials with a vertical stroke of red pigment. In choosing the initials to highlight in red, the rubricator seems to have semi-randomly picked visually prominent capitals in the recipes in addition to the first letters of the title and the recipe itself. The most common initials highlighted are large letters such as S, G and A. Also some punctuation symbols, especially at the ends of recipes, are similarly highlighted with a red stroke.<sup>8</sup>

## Later additions

Apart from the 18<sup>th</sup>-century foliation, there are very few later annotations or additions in the manuscript. The earliest of these would seem to be the addition of small symbols, resembling the typical form of a superscript *a* (Denholm-Young 1954: 67), into the left margin next to the titles of some recipes. The ink used for these annotations is of a very faint grayish brown colour (hand 2), and the closest match would seem to be the ink used for ruling the pages and for the original signatures. This is however somewhat unlikely, since unlike the ruling and the signatures, these annotations would seem to have been added after the text was written.

In addition to these roughly contemporary markings, there are some later notes added to the front and back of the text. On f. 1r there is an undated acquisition note by either Edward Harley, or David Casley, the deputy-keeper of the Cotton library who was curating his library at the time, stating that the manuscript was “Bought in mr Rawlinsons Sale Oxford BH S”, and on f. 28r there is a somewhat unclear closing note in the same hand, apparently reading “Finis huius Libri sibes | etate Finis huius Libri ibus | etate non veta libre”. The shelfmark “123.B.14” and the catalogue number “4016.” on the second flyleaf were most likely added by David Casley when he catalogued the manuscript in the Harley library sometime between 1734 and 1736. On f. 29r there is also a note by a British Museum librarian or cataloger dated 1885, reading “ff. 28 | 4o [3 illegible characters] Jan. | 1885”. The same hand has also added the formula “2/III e” onto the second line of the second flyleaf.

<sup>8</sup> This level of decoration would seem to be relatively typical for recipe collections, as for example Beinecke Library MS 163, which contains a copy of the *Ordinance of Pottage* collection and is described by Hieatt (1988: 9), is very similar in its presentation, being “not especially handsome, but [...] legibly written in a standard Anglicana hand of the period”, with headings placed centred on their own lines and underlined in red, and some initial capitals touched up in blue or red pigment.

### Origins and provenance

The latest date explicitly mentioned in the bills of fare is 1443, when John Stafford was made the Archbishop of Canterbury, which sets its a date of *ante que non*. In his edition, Austin (1888: vii) dates the manuscript to c. 1450, which is certainly a plausible suggestion. The relevant library catalogue (BL Harley: 104) contains no additional information about the origin of the text, the only characterization of the content being: “The whole in a very old hand writing.” The geographical origin of the manuscript, as well as its initial owner or scribe are unknown, but it seems to have been copied by a single scribe.

The manuscript contains no traces of its early history, but based on the fortunate inscription of f. 1r, the manuscript was apparently acquired, along with sixteen other manuscripts,<sup>9</sup> by Edward Harley, 2<sup>nd</sup> earl of Oxford and Mortimer, book collector and patron of the arts, in March 1734 when the manuscript collection of Thomas Rawlinson (1681-1725) was auctioned over sixteen days as one of the last parts of “the largest library that had been collected up to his time” (Wheatley 1888: 128; see also Wright and Wright 1966: lxxvii). The manuscript was catalogued sometime between 1734 and 1736 by David Casley, the deputy-keeper of the Cotton library.<sup>10</sup>

The manuscript was acquired by the British Museum in 1753, when it was bought from the daughter of Edward Harley, Margaret Cavendish Bentinck, Duchess of Portland for £10,000 (which was a fraction of their contemporary value) under the Act of Parliament that also established the British Museum (BL Help: Harley MSS). The manuscript was catalogued or otherwise processed—and apparently also foliated—in 1885 by a British Museum librarian, who inserted a note with this date in pencil onto f. 29r.

### 9.2.2 London, British Library MS Harley 279

#### Physical description

The manuscript is written on thin parchment of dark cream colour, with the grain pattern clearly visible on many of the pages. The leaves have not been trimmed to uniform size and the bound manuscript thus resembles a bundle of folia with rough edges. The manuscript is made up of 50 quarto or octavo sized parchment folia, 210-228 mm in height and 140-155 mm in width.<sup>11</sup> Ff. 1-5 and ff. 46-51 are slightly narrower than the rest, (c. 140-145 mm) and have been attached to the binding with modern paper strips, which make up their width to c. 150 mm, matching the rest of the manuscript. The six folia inserted to the end of the manuscript (ff. 46-51) are also slightly shorter than the rest (c. 210 mm). The thickness of the manuscript, excluding the covers, is c. 14 mm, varying slightly due to the covers being slightly curved, being thickest in the middle.

<sup>9</sup> The manuscripts acquired by Harley from this sale include Harley MSS 3862, 3961, 3991, 4012, 4015, 4016, 4107, 4108, 4120, 4124, 4135, 4137, 4635, 4690, 4692, 6426 and 7371.

<sup>10</sup> The handwritten catalogue, which forms volume VIII of the ‘*Catalogus Brevior*’, originally begun by Humfrey Wanley in 1708, is currently British Library Additional MS 45708. A printed version of it was published in 1759 by the British Museum that had acquired Harley’s books in 1753 (BL Harley).

<sup>11</sup> Gumbert (2000: 82) notes that the skins used in medieval Europe were usually between 44 and 80 cm in length, which means that the format could equally well be either a quarto cut from a small skin or an octavo cut from a larger one.

The parchment used for the manuscript is not of very high quality and has various kinds of imperfections. The bottom edge of f. 11 is quite irregular, with a triangular piece (c. 20 by 85 mm in size) missing from the outer bottom margin, most likely due to the sheet being taken from the edge of a hide which was slightly too small. The bottom margins of ff. 19v and 20r and the left lower corner of the text block on f. 19v show the follicles of the skin very clearly as small black dots grouped in threes, which would suggest that the parchment was made from pigskin instead of sheep or bovine skin. This pronounced surface texture also makes the text quite hard to decipher at the bottom of f. 19v. There is a semi-circular hole with radius of 10 mm at the bottom edge of f. 32 (foliated in MS as 31), its centre 65 mm from the inside edge and its edges slightly darker and thicker than the rest of the parchment. The preceding and following folia also show the outline of the hole as a slight darkening of the parchment within the edges of the hole.

In addition to imperfections that predate the writing of the text, the manuscript has also suffered damage during its later history. The first six folia (of which one is missing) have at some point become loose leaves and have been rebound (most likely in the 19<sup>th</sup> or 20<sup>th</sup> century) by attaching paper binding strips over their inside margins and sewing through them, leaving a paper stub opposite each folio. Judging from the contents, there seems to be a single folio missing from between ff. 4 and 5. The current order of the folia and stubs is as follows: f. 1, f. 2, f. 3, stub, stub, stub, f. 4, f. 5, stub, stub. Although the sewing is too tight to be seen, this arrangement means that these single leaves must have actually been sewn on as two groups.<sup>12</sup> Also ff. 46-51 have at some point become loose and have been fixed—perhaps at the same time as the first folia—by building them into bifolia with paper binding strips glued into the inner margins. They also seem to have at some point either suffered damage to their edges and been trimmed, or have originally been cut to a different size, since they are currently slightly smaller than the rest of the manuscript. Damage would seem a likely explanation, considering that many of these folia are severely stained.

The pages of the manuscript in general are quite dirty and stained, especially near the edges, but it is impossible to tell the source of the staining. Mostly this has merely darkened the pages slightly, not obscuring the text or diminishing its legibility. The edges of the pages have darkened more significantly to the depth of a few millimeters, mostly at the top and outside edges of the folia, which would suggest moisture as the cause. The condition of the parchment varies quite significantly, and despite the generally worn appearance, there are some pages which are of a quite clean cream colour, with just slight darkening around the edges. There is also some evidence of slight moisture damage along the bottom outer edge of the manuscript, resulting in slight discolouration of a roughly 1 by 4 cm area on the outer edge of the bottom margin.

There are also instances of more localized damage. On f. 48 (foliated in MS as 46), there is a diagonal tear, starting from the fold about 2/3 of the way up the page and extending down to about 50 mm from the bottom edge of the page and 40 mm from the fold, extending into the text block. Although the tear runs mostly in the

<sup>12</sup> For reasons of convenience and because they most likely originally constituted a single quire, these five single leaves have here been treated as a single quire.

inner margin and on line-fillers, there are some letters which have been rendered slightly unclear by it. There is also a dark brown liquid stain in the middle of the left half of f. 55v (foliated 50), which has also penetrated to ff. 52-54. The verso of the last parchment folio of the manuscript (f. 55v) is generally dirty and stained, possibly indicating that the manuscript has spent an extended period of time in an unbound state.

### Collation of the manuscript

In the current binding, the original manuscript quires are preceded by a single pasteboard and a modern flyleaf (with a note of printed editions of the texts in this MS) and followed by 2 modern flyleaves of different papers, the first of 18<sup>th</sup>-century paper with the chain lines clearly visible, and the second of smooth 20<sup>th</sup>-century paper (the same as in the beginning), along with a pasteboard. The collation of the manuscript, excluding the modern flyleaves, is as follows:

- 1<sup>6</sup>, consisting of single leaves pasted onto paper binding strips and bound with stubs, wants 5 (ff. 1r-5v);
- 2<sup>12</sup> (ff. 6r-17v);
- 3<sup>12</sup> (ff. 18r-29v);
- 4<sup>12</sup> (ff. 30r-41v);
- 5<sup>8</sup>, 3 bifolia (ff. 46-51)—consisting of six single leaves pasted together with paper strips—added to the middle of the quire (ff. 42r-55v).

**Signatures and catchwords** The single leaves making up the first quire (ff. 1-5) have no signatures. Ff. 6-11 (quire 2) are signed *a j*, *a ij*, *a iij*, *a iiij*, *a v* and *a vj*, while ff. 18-23 (quire 3) are signed *b j*, *b ij*, *b iij*, *b iiij*, *b v* and *b vj*. The signatures are added in a black ink that is darker than the brown one used to write the text on quire 2 but roughly the same colour as the ink used to write the text on quire 1. If the signatures are in the hand of the scribe writing the text, he most likely changed to a new batch of ink after writing quire 1 and signing quire 2 but before writing the text on quire 2. It is also possible, that the signatures could be in a different hand, as at least the *a* seems different from that used in the text. Ff. 30-35 (quire 4) are signed *c j*, *c ij*, *c iij*, *c iiij*, *c v* and *c vj*, in a hand similar to that used for the signatures of quires 2 and 3, but in red pigment. There are no signatures on quire 5 (ff. 42-55). There are catchphrases in the bottom margins of ff. 17v, 29v and 41v, i.e. on the last leaves of quires 2, 3 and 4, written in a hand that looks similar to the scribal hand used for the text itself but using a lighter brown ink. The catchphrase on f. 17v—“an whan þe”—is surrounded by a decorated box drawn in red pigment, while the catchphrase on f. 29v—“and put | on a potte . | þen putte”—has been written diagonally on three sections (separated here by vertical lines) of an angular scroll wrapped around a horizontal pole with decorative ends, the scroll emanating from and terminating in the open mouth of a man’s head looking up on the left and down on the right. All of this decoration is in the same ink as the catchphrase itself. The catchphrase on f. 41v—“of. Sugre þan colour.”—is placed within a narrow horizontal box, wrapped around with acanthus-like leafy scrollwork, all in the same brown ink. A later hand (which has also added the title page for the bills of fare on f. 46r) has added further catchwords to the bottom margins of ff. 45v-48v.

**Foliation** The manuscript has been foliated in pencil in the right top hand corner of each recto page, apart from ff. 27r, 38r, 52r, 53r and 54r, which have been ruled but not written on and therefore skipped by the foliator. The foliation has most likely been added by a librarian of the British Museum sometime during the 20<sup>th</sup> century.

### Manuscript layout

The layout is the same for the majority of the manuscript folia, consisting of a single column text block, outlined and lineated in hairlines of light brownish ink and measuring 155-160 mm in height and 100-105 mm in width. The remaining space is divided between the margins, the top margin measuring 20-25 mm, the bottom margin measuring 40-50 mm and the inner and outer margins measuring 12-20 mm and 30-40 mm, respectively. The pages have been pricked, and the pricks are still clearly visible at 2-10 mm from the outer edge of the page. Based on this, it would seem that the quires were pricked before being folded. The pricks are made with a square awl or some other implement that has left cross-shaped holes whose orientation and size varies. The lines outlining the text block—as well as the second and second-to-last lineations—extend to the edges of the page, while the rest of the lineations stop at the edge of the text block.

The text has been written under the top line, but any titles, headings and running heads have been written on top of the top line. The boundaries of the text block are mostly observed, although the text occasionally runs over the boundary at the right side of the block, but usually by no more than 5 mm. Each page has 29 ruled lines (not including running heads or titles above the top line). The height of the lines varies slightly, being just under 5.5 mm on average. The minimum height for the scribal hand is very consistently 2.5 mm or slightly less, which together with quite short ascenders and descenders gives the text block a quite clear and regular appearance. On the folia with the table of contents (ff. 1-5), there is an additional vertical ruling, subdividing a column of 15 mm from the right edge of the text block for the numbers of the recipes. The text blocks of the slightly smaller pages added to the middle of quire 5 (ff. 46-51), containing a series of bills of fare, have the same outside dimensions but are laid out in two columns, separated by a pair of vertical rulings down the middle of the page with a space of 7 mm between them.

In terms of marking textual organization, individual recipes are not separated from each other by empty lines, but each recipe begins on a new line with the title written in line with the rest of the recipe but in red pigment. Recipe numbers are inserted into the outer margin next to the title.<sup>13</sup> In the table of contents, laid out in a single column, the titles of the recipes are aligned to the left edge of the writing block, while the recipe numbers are aligned to the right edge, the intervening space being filled with line fillers consisting of curved strokes in red pigment.

<sup>13</sup> A similar visual layout of recipes—with each rubricated title beginning a new line with the recipe body following it on the same line, incomplete final lines filled by rubricated ornaments—can be found in a collection of mainly medical recipes in California Huntington Library MS 1336, for which a digital facsimile of a sample opening is provided by the Digital Scriptorium at <<http://digitalassets.lib.berkeley.edu/ds/ucb/images/heh/150/001488A.jpg>>(Carroll 2006: 311).

### Current binding

Judging from the markings on the binding, the manuscript has most likely been rebound into its current binding sometime after its sale to the British Museum in 1753. It was most likely in connection with this rebinding operation that the loose folia now making up quire 1 (ff. 1-5) were bound into their current configuration; it is unknown whether the folio missing from between ff. 4 and 5 was already missing at this point, or has been subsequently lost. The current binding is a half binding in purple imitation leather (or stiff cloth of similar appearance) with a black leather spine and front corners, the covers measuring 235 mm in height and 162 mm in width (excluding the spine which adds 5 mm to the width). The coat of arms of the Harley family—"Or, a bend cottised Sable"—supported by two angels, surmounted by the coronet of an Earl and accompanied by the motto "VIRTUTE ET FIDE" decorate the front and back covers. On the spine there are six cords surrounded by gilded lines dividing the spine into seven compartments into which the following text has been printed in gold (except for the shelfmarks at the ends of the spine, which are applied on white paper stickers): "50 - CULINARY RECIPES - BRIT. MUS. - HARLEY 279. - - - H.1".

### Summary contents of the manuscript<sup>14</sup>

- 1) A numbered list of recipe titles in three parts (ff. 1r-5r)
  - a) First part of a numbered list with 153 recipe titles (ff. 1r-3v)
  - b) Incomplete second part of a numbered list with 57 recipe titles (ff. 4r-4v)
  - c) Incomplete third part of a numbered list with 13 recipe titles (ff. 5r)
- 2) A three-part collection of 258 culinary recipes (ff. 5v-44v)
  - a) A sub-collection of 153 recipes titled *Potage dyvers* (ff. 5v-26r)
  - b) A second sub-collection of 64 recipes titled *leche vyaundez* (ff. 28v-37r)
  - c) A third sub-collection of 41 recipes titled *Here begynnyth dyuerse bake Metis* (ff. 38v-44v)
- 3) A collection of menus (ff. 45r-50r)
  - a) A title page for the collection of menus (f. 46r)
  - b) A menu of three courses served at the coronation of Henry IV (f. 47r)
  - c) A menu of three courses served at the second wedding of Henry IV (ff. 47r-v)
  - d) A menu of three fish courses served at the same wedding (ff. 47v-48r)
  - e) A menu of three courses for Trinity Sunday (f. 48r)
  - f) A menu of three courses served by Lord de la Grey (ff. 48r-48v)
  - g) A menu of three courses served by or for the bishop of Lincoln (f. 48v)
  - h) A menu of three courses served at the consecration of the bishop of Salisbury (ff. 48v-49r)
  - i) A menu of two courses served at the funeral of the bishop of Bath and Wells (ff. 49r-v)
  - j) A menu of two fish courses for the above funeral (f. 49v)
  - k) A menu of three courses served at the consecration of the new bishop of Bath and Wells (ff. 49v-50r)
  - l) A more modest menu of two courses served to the lesser guests at the same consecration (ff. 50r-v)
  - m) A menu of three courses served at the wedding of the Earl of Devonshire (f. 50v)
  - n) A more modest menu of two courses served at the same wedding (ff. 50v-51r)

<sup>14</sup> A more detailed description of the contents of this MS is contained in appendix F and in the TEI header of the base data file for the MS included in appendix A.

### Hands used in the manuscript

All of the items contained in the manuscript—except for the table of contents for the bills of fare (MS item 3a)—seem to be written in the same hand, although the colour of the ink varies throughout the manuscript. The bills of fare themselves are written in the same hand as the rest of the manuscript, as is the rubrication added to the text.

**Scribal hand** The original scribal hand, written in a black or dark brown ink, is a quite formal late example of *textualis semiquadrata*. The letterforms are typical to *textualis*—except perhaps the *a*, which is of the clearly two-compartment anglicana form instead of the straight-sided textualis form—and the minims have rounded heads and feet. The minim height is very consistently 2.5 mm or slightly less, which together with quite short ascenders and descenders gives the hand a quite regular appearance. The colour of the ink varies considerably throughout the manuscript, being a dark black in what is now the first quire (containing the tables of contents), and getting noticeably browner and lighter in the second quire. Since the second and third quires also look more worn—both in terms of the parchment and the ink—it is difficult to say whether the lighter colour is due to wear or to an original difference in the ink. The ink is especially light on ff. 18 and 19, after which point it gets darker, although not quite as black as on the first quire. From f. 28r onwards, the ink is again as black as it was in the first quire, until it again gets lighter at the end of f. 43v (foliated 41v), and remains a lightish brown to the end of the collection on f. 45v (foliated 43v). For the bills of fare appended to the collection, the ink is again a very dark brown, although it has been worn to a slightly lighter shade on ff. 47r-48r (foliated 45-46).

**Other hands** In addition to the scribal hand, the following hands can be identified in the manuscript:

- 1) The hand of the rubricator in red pigment that has added the titles and final line fillers to the recipes, framed the recipe numbers and highlighted the initials of significant words in red pigment. Since the hand is very similar to the hand in which the text itself is written—although slightly larger—the rubrication has most likely been done by the original scribe him- or herself.
- 2) A hand in black ink, possibly of anglicana type, that has added the signatures to quires 1, 2 and 3. The identification of the script is based solely on the shape of the letter *a*, which is an elongated two-compartment one typical to anglicana. Based on the colour of the ink, which is similar to that used for the text on quire 1, the hand could conceivably belong to the original scribe, but since the shape of the *a* is somewhat different from that in the text, it could very well be by someone else.
- 3) A hand in red pigment, possibly of anglicana type, that has added the signatures to quire 4. The hand would seem to be the same as that used for the signatures in quires 1, 2 and 3, except for the medium. If the hand belongs to the original scribe, the red pigment could be explained by the scribe signing the previous quires at the same time he wrote the text itself, but signing the fourth quire later when he added the rubrications.
- 4) A textualis hand in light brown ink that has added the catchphrases to ff. 17v, 29v (foliated 28v) and 41v (foliated 39v) and some of the decoration surrounding them. Based on its letterforms and general aspect, the hand could very well belong to the original scribe, but the difference in the colour of the ink—which is a much lighter brown than the text even at its lightest—suggests that the catchwords were written

separately from the text itself.

- 5) An unknown hand in red pigment, different from that used by the main rubricator, that has added the red decoration surrounding the catchphrases on ff. 17v and 29v (foliated 28v). The fact that this red pigment is different from the one used by the main rubricator would seem to support the later addition of the catchphrases themselves, either by the original scribe or by a later annotator.
- 6) A formal textura hand in black ink that has added the phrase “*Tertria sunt*” at the beginning of the first ruled line of f. 55v (foliated 50). This hand could belong to the original scribe, but since it is larger and more formal, it could just as well be either a later addition or the aborted start of another text (perhaps on algebra), for which the parchment sheet in question was originally intended.
- 7) An early or late modern italic hand in brownish gray ink that has added arabic numbers to the right of the original recipe numbers for some of the recipe titles in the table of contents. Since the hand has only written arabic numerals, the identification and dating of the hand are very uncertain. Since the original numbering in Latin numerals is very clear, it is likely that these numbers were added by someone unused to roman numerals, probably at a much later date.
- 8) An early or late modern italic hand in brownish-black ink that has added a library shelfmark (37.A.18) in the top margin of f. 1r of the manuscript, as well as the arabic numerals 1, 2 and 3 to the right of the first three recipe titles on the same page. Since the shelfmark which is consistent with the shelving system of Robert Harley’s library, the hand is likely to belong to his librarian Humfrey Wanley.
- 9) A 17<sup>th</sup>-century italic hand in greyish black ink that has added catchwords in the bottom margin of ff. 1v, 2v and 3v, possibly at a stage when the pages in question had either come loose or were in the danger of doing so. Judging from the comments of Watson (1966: 51), these catchwords might have been added by Sir Simonds D’Ewes sometime in the early 17<sup>th</sup> century. The hand seems, however, different from hand 10 (associated with D’Ewes), but this might be simply the result of the obvious imitation of gothic angularity in the latter hand.
- 10) A large angular italic hand in grayish black ink, used to add a title page (with a table of contents) for the list of bills of fare on a blank (but ruled) page (f. 46r, foliated 44) and to add catchwords to the bottom margins of ff. 45v-48v. This hand, which is very calligraphic and decorative, giving the impression of an attempt to emulate the features of a medieval textualis hand in an italic context, most likely belongs to the 17<sup>th</sup>-century owner of the book, Sir Simonds D’Ewes, who had a habit of adding things to the manuscripts he owned (Watson 1966: 50-1, 212).
- 11) A modern italic hand in pencil—most likely belonging to the British Museum librarian who catalogued or otherwise processed the manuscript in 1875—used to add the foliation to the top right corner of each recto. The same hand has also added the current shelfmark (279) and the marking III<sup>2</sup>B in the top margin of f. 1r, and the notes “f. 50” and “Ex<sup>n</sup> Irr. July 1875” on f. 56r.
- 12) An unknown hand in black ink, which has added a line of dots to the right of the recipe titles on f. 3r in the table of contents.

## Decoration

The manuscript contains more decoration than most manuscripts of utilitarian nature, in line with its relatively formal hand and regular layout. This decoration clearly serves a text-organising function and consists mainly of the use of rubrication and large decorated initials to mark textual divisions such as individual recipes and subgroups within the collection.

The titles of all the recipes—as well as the running titles in the top margin—have been highlighted by writing them in red, and the end of each recipe has been marked by filling the empty space on its last line with an ornamental line filler



consisting of a series of wavy lines interlinked to form a braid-like pattern. The same line filler pattern has also been used to fill up the space between the title and number of a recipe in the tables of contents on ff. 1-5. The recipe numbers found in the margins of the collection itself next to each recipe title have been decorated by framing them on three sides (sides and bottom) with red lines. Important words, such as the names of dishes in the table of contents or names of ingredients in the recipes themselves, have been highlighted by picking out their initial letter with a vertical dash of red pigment.

Individual large decorated initials occasionally occur in the text at the beginnings of major text divisions. The first item in the table of contents on f. 1r, “Lange wortys de chare” has a rubricated Uncial-style initial of a single-line height, as does the initial of the same title on f. 6r in the recipe collection itself. Some running titles and headings located on the first line of a page have tall decorative ascenders, terminating in flourishes of varying elaboration. The initial *T* of recipe 129 on the first line of f. 23r—although similar in shape and ductus to all the other capital *T*s in the text—is roughly three lines in height and has been elaborated into a profile portrait of a man wearing a pointed hat. The initial *T* of the first recipe in the second part of the recipe collection on f. 27v is similar in size and shape, but without the portrait. A decorative looped approach stroke on the initial *V* of the running title “Vyaunde leche” on f. 37r has a frontal view of a face drawn in red within it.

The catchphrases on ff. 17v, 29v and 41v—which seem to be a later addition—have been decorated in various ways. The one located in the middle of the bottom margin of f. 17v is surrounded by a rather crude scroll-like frame with floriated decorations emanating from its corners, all drawn in the same light brown ink in which the catchphrase is written. The catchphrase on f. 29v has actually been integrated into a decorative illustration that fills up the entire bottom margin. The illustration consists of a round horizontal column—drawn in the same ink as the catchphrase itself and shaded to look three-dimensional—with decorative finials at each end, around which is wrapped a crudely drawn and quite angular ribbon, drawn in red and emanating at both its ends from the mouths of two profile portraits of male heads, which are again drawn in the same light brown ink as the catchphrase. The catchphrase itself has been written on the three segments of the scroll that cross over the column. The catchphrase on f. 41v is written on a straight horizontal bar, around which is wrapped an acanthus-shaped ribbon or vine, dividing the bar into four segments.

### Later additions

There are very few annotations or corrections made to the manuscript by later users of the text, which may be due to the unusually high production values of the manuscript (i.e. book hand written on parchment). Judging from the ink, the few corrections that are found in the text seem to be made by the original scribe himself, although it is difficult to compare the hands since the interlineal additions are in a very small and cramped hand. Most of these corrections have clearly been made already when writing, since they are inserted in-line immediately after the erasure of the erroneous form.

Two sequences of partial numbering in arabic numerals have been added to the table of contents on ff. 1-5. Although neither can be dated with any certainty,

the one added immediately to the right of the original recipe numbers in hand 5 would seem to be earlier than the ones added to the outer edge of the margin in hand 8, since the number forms used by it are more archaic in shape (especially the laterally symmetric number 4 with a closed loop). The earlier sequence consists of numbers 1-29 on f. 1r, numbers 1-20 and 1-9 on f. 2r (of which all but 1, 2, 4 and 6 in the first series and 4 and 9 in the second series have been crossed out in the same hand). There might also have been similar numbering on f. 1v, but if there was, it would be obscured by the strip of paper used to reattach the page to the binding. Instead, ff. 1v and 2v contain small points in the same ink next to each recipe, probably counting off the recipes; on f. 2v, the points are visible only from recipe 104 onwards and not all recipes have the points (which are included in the transcription).

There are traces of a coat of arms in the inner bottom corner of the last parchment folio, sized c. 40 x 40 mm. It is difficult to tell on which side of the folio the coat of arms has originally been drawn, since the drawing has penetrated the page and has subsequently been washed away almost completely, leaving only faint traces on both sides. The coat of arms would seem to be “argent, three crescent moons sable”, but since the drawing seems to have been monochrome, the metal and colour are indeterminate, and could even be the other way around. Because of this it is difficult—if not impossible—to determine with certainty the owner of this device, although if added by Sir Simonds D’Ewes, it could belong to the D’Ewes family.<sup>15</sup> An 18<sup>th</sup>-century paper flyleaf (f. 56) has been bound to the end of the manuscript with two somewhat unclear notes written in pencil (hand 11) on its recto side, the first reading “ff. 50” and the second (possibly) “Ex<sup>n</sup> Irr. July 1875”.

### Origins and provenance

Since the latest date explicitly mentioned in the bills of fare accompanying the recipe collection is the 16<sup>th</sup> of September 1425, at least that part of the manuscript must have been copied after this date. In light of this date, the reference to the wedding of the Earl of Devonshire most likely refers to the wedding of Thomas Courtenay, the 13<sup>th</sup> Earl of Devon and Margaret Beaufort, which most likely took place sometime between 1430 and 1432 (Cherry 2004). This would push the *terminus post quem* to the early 1430s. This is also the dating given by Austin (1888). The geographical origin of the manuscript, as well as its initial owner or scribe are unknown, but it seems to have been copied by a single scribe.

Due to the lack of early inscriptions of ownership or other signs of use, the early history of this manuscript is unknown. The first owner we know something about is Sir Simonds D’Ewes (1602-1650), diarist, antiquary and a lover of historical records, who acquired the manuscript from an unknown source sometime between 1620 and 1650 (Watson 1966). Upon the death of Sir Simonds D’Ewes in 1650, his collection—along with the estate and title of baronet that he had acquired in 1641—passed on to his only son (by his second wife, Elizabeth Willoughby, whom he had married in 1641 after the death of his first wife Anne Clopton), Sir Willoughby D’Ewes (c. 1650-1685), with the stipulation that it should be accessible

<sup>15</sup> Unfortunately circumstances did not allow me to consult BL MS Harley 381, which, according to BL Harley, contains several documents relating to the genealogy and history of the D’Ewes family, but this source will be examined upon a future visit to the library.

to “all lovers of learning” (Watson 1966: 54).

In 1705, on the 4<sup>th</sup> of October the collection of Sir Simonds D’Ewes, first baronet D’Ewes, including this manuscript, was bought by Robert Harley from his grandson, also named Sir Simonds D’Ewes, for £450 to form the foundation of his manuscript library. The purchase of the collection, which consisted of over 600 manuscripts in addition to printed books, prints, drawings and coins, was negotiated by his librarian Humfrey Wanley, who also catalogued the manuscripts three years later.<sup>16</sup> (Heyworth 1989: 232-4; Wright and Wright 1966: xviii-xix.) The manuscript was acquired by the British Museum in 1753, when it was bought from the daughter of Edward Harley, Margaret Cavendish Bentinck, Duchess of Portland for £10,000 (which was a fraction of their contemporary value) under the Act of Parliament that also established the British Museum (BL Help: Harley MSS). The manuscript was catalogued or otherwise processed—and apparently also foliated—in 1875 by a British Museum librarian, who inserted a note with this date in pencil onto f. 56r.

### 9.2.3 London, British Library MS Additional 5467

#### Physical description

The manuscript is a composite one written on two different stocks of paper, which have by now acquired a brownish grey colour. At some point of their history, the edges of the manuscript pages have suffered damage to the extent that they have been restored by leaf-casting (Futernick 1983) new edges of noticeably lighter paper around the original folia to make up for the lost margins. After recasting, the pages have been retrimmed and their corners rounded to a radius of c. 20 mm. The watermarks on the two paper stocks have been identified and are described by the BL MSS Catalogue as follows:

- A) (ff. 1–22) Tête de boeuf, Briquet (1923) nr. 14183.
- B) (ff. 23–224) Circle with linear design.

In terms of its structure, the manuscript is composed of three modern flyleaves and a total of 224 original folia (and a single folio fragment). The folia are of a typical Chancery quarto size (Gumbert 2000: 81-2), measuring approximately 210 mm by 140–145 mm in their current state, including the restored margins, with the laid lines running vertically. This is most likely slightly—but not much—smaller than the original size of the pages, judging from the fact that no marginalia has been obviously cropped but some of it is currently located quite close to the edge. The thickness of the manuscript, excluding the covers, varies between 55 and 60 mm, being largest in the middle. The manuscript is missing a part at the end, as indicated by the catchword on the verso of the last folio (f. 224). Also the remaining leaves of the manuscript have been badly damaged at some point, resulting not only in the loss of a large portion of the margins, but also in the separation of bifolia into individual leaves. This means that the current collation of the pages might not reflect the original one, as individual leaves have been rejoined into bifolia by a span of new paper, through which they have been sown into

<sup>16</sup> This part of the manuscript collection is catalogued in British Library Additional MSS 45701 and 45702. A printed version of this ‘*Catalogus Brevior*’ was published in 1759 by the British Museum (BL Harley).

the current quires. Somewhat puzzlingly, the BL MSS Catalogue states that the margins are “Contemporary - 17<sup>th</sup> century”, implying that the date of the recasting process is unknown to the cataloguer and was therefore not performed at the British Library, but earlier in the life of the manuscript. Based on the fact that the foliation in arabic numbers has in places been truncated by the damage to the edge or covered by a fine layer of paper pulp, the recasting (and perhaps some of the damage) must have taken place after the pages were foliated.

There is also some moisture damage to the remaining original page edges, occasionally extending past the edge of the text block. The first part of the manuscript (ff. 1–22) (especially the first two folia) has been more severely affected, to the point of occasionally hindering legibility at the beginnings and ends of lines at the outer edge of the page, especially on the top line where the damage is heaviest. It would seem that either this part of the manuscript suffered water damage while it was not yet attached to the rest of the manuscript or the ink used for it was more susceptible to water damage than the one used elsewhere. In the rest of the manuscript, the moisture has occasionally caused the ink to fade slightly but does not seriously affect its readability. The water damage seems to have occurred before the foliation in arabic numbers, as this does not seem to be affected by moisture. There is a largish diagonal stain of translucent brown colour on ff. 51 and 52, which has also penetrated all the way to f. 49 and to f. 55. Judging from the efficiency with which the stain has penetrated the paper, it would seem to be of a greasy nature, but it has not made the paper translucent, as would be expected of a grease or oil based substance. The existence of this stain alone does not tell us much of the use of the manuscript, but does hint at the possibility of the manuscript having at some point being used outside of a clean library environment.

### Collation of the manuscript

The collation of the manuscript, excluding the three modern flyleaves and divided into its component parts, is as follows:

A. ff. 1–22:

- 1<sup>10</sup> (ff. 1r-10v);
- 2<sup>10</sup>, 2 single leaves (ff. 11–12) added (ff. 11r-22v).

B. ff. 23–224:

- 3<sup>12</sup> (ff. 23r-34v);
- 4<sup>12</sup> (ff. 35r-46v);
- 5<sup>12</sup> (ff. 47r-58v);
- 6<sup>10</sup> (ff. 59r-68v);
- 7<sup>12</sup>, wants almost all of 5, fragment of it remains between f. 72 and f. 73 (ff. 69r-79v);
- 8<sup>10</sup>, 1 single leaf (f. 85) added (ff. 80r-90v);
- 9<sup>12</sup> (ff. 91r-102v);
- 10<sup>12</sup> (ff. 103r-114v);
- 11<sup>12</sup> (ff. 115r-126v);
- 12<sup>10</sup>, 1 single leaf (f. 133) added (ff. 127r-137v);
- 13<sup>12</sup> (ff. 138r-149v);
- 14<sup>12</sup> (ff. 150r-161v);
- 15<sup>12</sup> (ff. 162r-173v);
- 16<sup>12</sup> (ff. 174r-185v);

- 17<sup>12</sup> (ff. 186r-197v);  
 18<sup>12</sup> (ff. 198r-209v);  
 19<sup>12</sup> (ff. 210r-221v);  
 20<sup>2</sup>, 1 single leaf (f. 223) added (ff. 222r-224v).

**Signatures and catchwords** Ff. 47–48, 51 and 53–54 (quire no. 5) have signatures *a-c* and *e-f* in what seems to be the original scribal hand at the top left corner of the recto side. Additionally, f. 46v (quire no. 4) has the signature or catchword *b* at the bottom right corner of the page, the meaning or function of which is unknown. While the damage to the margins explains the missing signature *d* (which was most likely on f. 52r), the lack of a signature on ff. 50–51 is not explained by a similar gap in the sequence. The irregularity in the signatures could be seen as a sign of confusion in the order of the folia (made more likely by their modern rebinding), but the text of MS item 3 seems to run as intended even at the point of the problematic signatures. The most plausible explanation for this anomaly is a mistake in the original signing of the leaves, the scribe turning over three leaves (ff. 48–50) at once when inserting the signatures, most likely before writing the text.

The recto sides of ff. 69–74 (quire 7) have signatures *a-d* and *f-g* and ff. 80–85 (quire 8) have signatures *a-f* in the top left hand corners of the page. The verso sides of ff. 91–95 (quire 9) have signatures *a-e* on the right side of the bottom margin, the *f* being found in the same place on the recto off. 96. The verso sides of ff. 103–108 (quire 10), ff. 115–120 (quire 11) and ff. 127–132 (quire 12) also have signatures *a-f* on the right side of the bottom margin. Ff. 138–142 (quire 13), have similar signatures, but only up to *e*, the *f* having either been lost due to damage or omitted by the scribe. Ff. 150–156 (quire 14) have a similar series from *a* to *f*, with *c* and *d* having been lost. Ff. 162–167 (quire 15), ff. 174–179 (quire 16), ff. 186–191 (quire 17) and ff. 198–203 (quire 18) again have the full series of signatures (*a-f*) in the same place, while ff. 210–215 (quire 19) have the same series but for *c* and *e*, which have been lost due to damage to the edge of the manuscript. Ff. 222–224 (quire 20, incomplete) have signatures *a-c* in the same place. All of the signatures are apparently in the original scribal hand.

Catchwords are found at the bottom of ff. 11v (currently the first folio of quire 2), 72v and 73v (currently the third and fourth leaves of quire 7), 85v (a single leaf added to the middle of quire 8), 86v, 87v, 88v, 89v and 90v (currently the last 5 leaves of quire 8). In quire 9 (ff. 91–102, all of the verso sides have a catchword in the right part of the bottom margin, except for f. 96, where the catchword is in the same place on the recto side. In the latter half of the manuscript, most of the leaves have a catchword in the original hand in the bottom margin of their verso sides, apart from ff. 147, 156–157, 166–172, 180–185, 188–189, 192–196, 204–208 and 216–220.

**Foliation** Original medieval foliation in an unknown hand (hand 5) using Roman numerals is mostly preserved up to f. 96 (which is numbered *c*), although some of the intervening foliations have been lost to damage. The original foliation probably continued throughout the manuscript; for example f. 139 still bears the original foliation *cxlv*. This means that the two parts of the MS were bound

together already in the medieval period. A later italic hand (hand 6) has foliated the manuscript throughout in arabic numerals and black ink at the top right corners of the rectos. This was apparently done before the pages were damaged and certainly before they were recast (see below), as some of the folio numbers have been truncated by the damage or covered by a thin layer of the paper pulp, which has formed a translucent white film on top of the ink.

### Manuscript layout

The layout is very similar throughout the whole manuscript, consisting of a single column text block outlined in dry point, measuring 145–155 mm in height and 80–95 mm in width. The ruling seems to be stronger in the beginning of the manuscript and get fainter towards the end, to the point of being almost invisible. The remaining space is divided between the four margins, the top margin measuring 15–25 mm, the bottom margin 35–45 mm and the inner and outer margins 15–20 mm and 35–40 mm, respectively.<sup>17</sup> There are no signs of pricking in the manuscript, which explains the fluctuation in the size of the writing block. No lines have been ruled, and the boundaries of the text block are observed to a varying degree, the first line of text being frequently written above the top line and the lines of text often crossing into the right-hand margin, giving the pages a rather irregular and loosely arranged appearance. An exception to this layout is formed by ff. 23–24, which contain a list of recipes and are ruled for two columns of writing by two vertical rules marking the left edges of the columns, located c. 65 and 130 mm from the outer edge of the page on the recto sides and c. 20 and 75 mm from the outer edge on the verso pages. The recipe titles begin at this line, the right edge of the column being ragged and depending on the length of the individual titles. This layout corresponds to that described by (Boffey and Thompson 1989: 286) for the anthologies compiled and copied by John Shirley, which are written mostly on paper, with a “practised, but not in any way refined” layout of single columns within a framework ruling.

The line height varies around 5 millimeters, with considerably larger spaces (or even empty lines) surrounding titles and headings. Since the average minimum height of the text is around 1.5 mm, varying between 1 and 2 mm, the lines of text are quite well-spaced and provide ample space for ascenders and descenders. Since individual lines of text have not been ruled, the number of lines per page varies, most pages having 30 or slightly fewer lines, depending on the number of headings or titles on the page, which take up the space of 3 to 4 lines of running text. In terms of marking textual organization, each recipe is divided from the preceding one by empty space, roughly one to three lines in height, containing the recipe number at the left edge and the recipe title more or less in the center.

### Current binding

The manuscript has been rebound into its current binding in the British Library in 1982. The rebinding involved stitching the bifolia—which had apparently been reconstituted earlier, as the BL MSS Catalogue seems uncertain as to the date of

<sup>17</sup> Except for quires 15–18 (ff. 162–209), where the bottom margin is about 10mm taller and the text block correspondingly shorter.

the operation—to 25 mm wide extenders made of thick, folded paper to facilitate opening the manuscript without damaging the brittle fold. Because of this, the current collation is very easy to make out, but does not necessarily conform to the original one. The binding itself is a half binding in brick red cloth with burgundy leather for the spine and corners, the covers measuring 225 mm in height and 170 mm in width (excluding the spine which adds 20 mm to the width). The thickness of the whole binding, including the boards varies between 65 and 69 mm due to significant convex curvature in the boards. On the spine, there are 5 raised bands, with the following text printed in gold, except for the shelfmarks at the ends of the spine, which are applied on white paper stickers: “117 - MISCELLANEOUS TRACTS - BRITISH LIBRARY - ADDITIONAL MS. 5467 - - A.13”.

### Summary contents of the manuscript<sup>18</sup>

#### A

- 1) Treatise on the tilling of trees (ff. 1r-15v)
- 2) Treatise on planting and grafting trees (ff. 16r-21r)

#### B

- 3) A Middle English collection of culinary recipes preceded and followed by a table of contents and bills of fare (ff. 23r-66v)
  - a) A list of recipe titles (f. 23r)
  - b) Five menus for royal banquets in Middle English, with some French and Latin (ff. 23v-25r)
    - i) A bill of fare for three courses (f. 23v)
    - ii) A second menu of three courses (ff. 23v-24r)
    - iii) A third menu of three courses (ff. 24r-v)
    - iv) A fourth menu of three courses (f. 24v)
    - v) A fifth menu of three courses (ff. 24v-25r)
  - c) A five-part collection of 181 Middle English culinary recipes with a main part of 91 recipes and four sub-collections (ff. 25r-64r)
    - i) Sub-collection of 14 recipes, titled *Diuerse sauces pour diuerse vyaundes* (ff. 45v-48r)
    - ii) Sub-collection of 20 recipes titled *La manere pur roster et saucer diuerse viaundes et cetera* (ff. 48r-51r)
    - iii) Sub-collection of 22 recipes titled *La maner pur roster buler et frier diuerse pessones* (ff. 51r-55r)
    - iv) Sub-collection of 34 recipes titled *Dyuerse Viaundes In Quaresme* (ff. 55r-64r)
  - d) A list of 184 recipe titles (ff. 65r-66v)
- 4) *Stans puer ad mensam* (ff. 67r-68v)
- 5) Medical and astrological advice (ff. 69r-72r)
  - a) List of unlucky days in each month in Middle English (f. 69r)
  - b) List of unlucky days for certain activities in Middle English (f. 69v)
  - c) List of lucky days for bloodletting in Middle English (f. 69v)
  - d) Prognostics from the weekday of Christmas in Middle English (ff. 70r-71r)
  - e) A prophecy attributed to Merlin in Latin (f. 71r)
  - f) Diet and regimen for different months in Middle English (f. 71r-v)
  - g) List of inauspicious days for bloodletting in Middle English (ff. 71v-72r)
  - h) Incomplete list of perilous days (?) in Middle English (f. 72r)
- 6) A Middle English account of the murder of King James I of Scotland in 1437, translated from a Latin version by John Shirley (ff. 72v-84v)

<sup>18</sup> A more detailed description of the contents of this MS is contained in appendix F and in the TEI header of the base data file for the MS included in appendix A.

- 7) A short treatise on pestilence in Middle English (ff. 85r-87r)
- 8) The treaty of Canterbury between Henry V and Emperor Sigismund in Middle English (ff. 87v-96v)
- 9) John Shirley's Middle English translation of the Book of Good Manners by Jacques le Grand (ff. 97r-211r)
- 10) John Shirley's Middle English translation of the *Secretum Secretorum* (ff. 211r-224v)

### Hands used in the manuscript

The different items contained in this miscellany manuscript are all written in one or several very similar secretary hands with strong Anglicana influence (or secretary-Anglicana hybrids) that include characteristically Anglicana letterforms like the *d* with a looped ascender, the 'sigma-shaped' *s* and the looped Anglicana *w*. The BL MSS Catalogue suggests the manuscript may have been written by two main scribes, one being responsible for the first part of the manuscript (ff. 1-21), and the second—"possibly 'Rouland Brugge'" or RoulandBrugges, whose name occurs on ff. 139 and 211—for the rest. This is possible, since the hands used for MS items 3-10 are very similar (apart from some variation in MS items 9 and 10, see hand 4), although the hand used for MS items 1-2 could equally well belong to the same scribe. Subsequent annotations, foliations and other markings have been added to the manuscript by various other hands, not all of which can be identified. The notes below document all of the hands occurring in the edited recipe collection, and those hands that can be associated with a name.

**Scribal hand** The original scribal hand used for the recipe collection—as well as for the medical advice (MS item 5) and for the text on the death of king James I (MS item 6) is a very small and compact secretary-anglicana hybrid with a minimum height of 1-2 mm. The hand is semi-cursive in shape yet quite regular and readable, becoming more cursive towards the end of the text. The hand frequently features long horizontal strokes trailing off at the ends of letterforms, which become more common as the text becomes more cursive towards the end of the text. Written in dark, somewhat cold brown ink, that has gotten lighter and warmer brown in areas affected by water damage.

In terms of specific letterforms, the hand is an interesting hybrid of secretary and Anglicana features. Both the single-compartment secretary *a* and the two-compartment Anglicana one are used, the former being the predominant one in the initial part of the text, roughly up to f. 32r, from which point on the two-compartment form gradually becomes more common, achieving clear prominence by f. 33v, and being almost exclusive for the latter part of the text. A similar progression is visible in the case of *g*, which is exclusively of the open-tailed secretary form in the beginning of the text up to f. 28r, after which the 8-shaped Anglicana form begins to appear alongside it and becomes the exclusively used form by the end of f. 29r. The *d* is of the looped Anglicana form throughout and word-final *s* occurs exclusively in the sigma-shaped Anglicana form, while *r* takes the short secretary form. The *w* occurs in both forms, both forms often occurring in close proximity (as in the phrase "draw it thorow" halfway through f. 27v, where "draw" has the Anglicana form while "thorow" has the secretary one. Also the *e* is found in both the 'reversed' Anglicana form and the 'traditional' lobed secretary form,



the latter being the predominant form throughout the text.

Some recipe titles in the recipe collection—as well as titles in the text on the death of James I—exhibit a more formal version of this scribal hand that undoubtedly belongs to the same scribe who wrote the body text (possibly Rouland Brugge). It can be characterized as a larger and quite formal textura-like display script written with a rather wide nib, but with dimensions more typical of secretary, including long ascenders and descenders. The overall aspect is very angular, even prickly, and minims terminate in diagonally elongated diamond-shaped serifs. Written in the same dark, somewhat cold brown ink, that was used for the body text and most likely written at the same time with it.

**Other hands** In addition to the scribal hand used for the recipe collection, the following hands can be identified in the manuscript:

- 1) The hand used for the two treatises on trees (MS items 1 and 2), a quite small and cursive anglicana which exhibits most of the characteristic anglicana letterforms, like the two-compartment *a*, the *d* with a looped ascender, the *f* and long *s* that have a long tail descending far below the baseline, the '8-shaped' two-compartment *g*, the sigma-shaped *s* the reversed anglicana *e*, and the large double-looped *w*. The most prominent secretary feature is the short *r* that does not descend below the baseline, which is used (along with the occasional '2-shaped' *r*) instead of the long *r* typical to anglicana. Another secretary letterform that occurs quite frequently is the pointed secretary *e*, which is used alongside the anglicana form in all positions. The hand is written in a light brown ink that has suffered significantly from water damage.
- 2) The scribal hand used for Lydgate's *Stans puer ad mensam* (MS item 4), the treatise on pestilence (MS item 7) and the treaty (MS item 8). In terms of both letterforms and the overall appearance, it is very similar to the hand used for MS items 1 and 2 (see hand 1), and slightly more cursive but otherwise similar to the hand used for the recipe collection (MS item 3) and tentatively ascribed to Rouland Brugge. It is, however written in an ink that is lighter, redder and 'warmer' in colour.
- 3) The very formal and decorative textura semiquadrata hand with frequent decorative ascenders and descenders, used for the explicit to the text on the death of king James I (MS item 6) and the incipits, explicits running heads and section titles found in MS items 4, 7 and 8. It is written in the same ink and clearly by the same scribe as hand 2 above.
- 4) The scribal hand used for the *Livre de Bones Meurs* (MS item 9) and the *Secretum Secretorum* (MS item 10). It is very similar to the other hands attributed to the second scribe, at least for the beginning of MS item 9, but its aspect and the ink used seem to vacillate several times over the two texts, sometimes changing characteristics several times within a single page. This points towards MS items 3–10 all being in the hand of a single scribe, whose hand occasionally alters either purposefully, or due to weariness or some other such reason. Around f. 105, the width of the pen nib seems to increase, along with the size of the script, while the colour of the ink and the letterforms remain similar. This increase in size affects only a few pages, and the nib width and script size return to their original size around f. 108. The colour of the ink seems to gradually grow towards a darker and colder shade, and by f. 164 it is already very similar to the very dark brown ink used for the recipe collection (MS item 3). Ff. 175–178 again seem to be written with a wider nib and a slightly larger hand. Folio 195 sees the hand change quite noticeably about 4 times, and carry on with a slightly different and more diagonal aspect from that point on. From this point on, the prominence of secretary letterforms, such as the straight-backed *d* and the simpler unlooped *w*, also increases. Apart from some enlargement of the script again on ff. 199–200, this same hand, still characterized by strong and straight diagonal ascenders slanting to the right, continues to the end of the manuscript, getting noticeably darker for the *Secretum Secretorum* text (MS item 10). This hand (and pos-

sibly also the original scribal hand for the recipe collection, as well as hand 2, hand 3 and hand 5) might belong to *Rouland Brugge* or 'Rouland Brugges', whose name occurs on ff. 139 and 211—for the rest.

- 5) A medieval secretary-anglicana hybrid hand in brown ink that has been used for foliating the pages in roman numerals. The hand is very similar to the hands used to write part B of the manuscript and could belong to the same scribe, espacially since the letterforms are similar and the ink used is of similar colour and has reacted to moisture in much the same way.
- 6) A modern hand in brownish black ink used for foliating the pages in arabic numbers. The time of the foliation is unknown.
- 7) A hand, in very black ink, that could be either medieval or modern, used for marking recipe titles with dots or asterisks. Since the symbols drawn in this hand are not script-specific, the script of this hand remains unknown, and it is difficult to tell to which of the manuscript's owner's this hand belongs.
- 8) An italic hand in dark black ink, presumably belonging to Marmaduke Towlard and used for marginal scribbles throughout the manuscript, including signature trials for the name "Marmaduke Towlard" and the name "John Towlard".
- 9) An early modern secretary hand in black ink that has written a marginal scribble in the lower margin of f. 43r.
- 10) A late 15<sup>th</sup>- or early 16<sup>th</sup>-century secretary hand that has written the name *Henry Smythe of Gerlthorpe* on f. 208v in black ink.
- 11) A late 15<sup>th</sup>- or early 16<sup>th</sup>-century secretary hand that has written the names *Henry Smythe* and *Wyllam Lyddall* in faded brown ink on f. 22v.
- 12) An early 16<sup>th</sup>-century secretary hand, written with a rather wide nib and black ink, that has written the name *Thomas Roberteson* on f. 68v.
- 13) A decorative 16<sup>th</sup>- to 17<sup>th</sup>-century hybrid secretary hand, written with a wide nib and brown ink, that has written the inscription "John Hall have deluud vnto Thomas Hall" on f. 106v and most likely belongs to one of the mentioned persons.
- 14) A 17<sup>th</sup>-century italic hand in purplish black ink that has written the name *William Towlard* on f. 158r.
- 15) A 17<sup>th</sup>-century italic hand in black ink that has written the name *John Thornton* on f. 96v.
- 16) A decorative 17<sup>th</sup>-century italic hand in reddish brown ink that has written the name *Richard Atkinson* on f. 97v.
- 17) A 17<sup>th</sup>-century italic hand in purplish black ink that has written the name *William Stowcroft* on f. 158r.
- 18) A 17<sup>th</sup>-century italic hand in dark brown ink that has written the name *Ann Lee* on f. 201v.

## Decoration

There is no rubrication or any other coloured decoration anywhere in the manuscript. The only form of decoration are some quite intricate initials drawn in ink using curving and floriate lines. This is very similar to the decoration described by for the anthologies compiled and copied by John Shirley which are ornamented by himself by penwork flourishes on major initials, and extended and flourished ascenders and descenders on the first and last line on the page.

## Later additions

Three modern flyleaves have been added to the front of the manuscript in connection with its 1982 rebinding. The first of these flyleaves carries the following note:

## DEPARTMENT OF MANUSCRIPTS

Record of Treatment, Extraction, Repair, etc.  
of MS. no. Add 5467

Date	Particulars	Name
25-2-82	Examined after binding.	T.A

Some longer annotations, some of which could be considered MS items in their own right (and have been so considered by Matheson (1999: 8–9)), have been added onto partly or entirely empty pages in the manuscript in later hands. These include:

- a) moral sayings on ff. 117v, 146v, 149v, inserted in a hand that has also added the signature *H. P.*, dated 24 June 1597, on f. 149v;
- b) moral verses on f. 196r, along with a signature and note of ownership by Edward Parker, dated to the late 16<sup>th</sup> or early 17<sup>th</sup> century by the BL MSS Catalogue; and
- c) religious and moral memoranda on ff. 21–21v, Latin verses, signature and names on ff. 22, 63v, names of family members on f. 82, accounts, etc. on ff. 155v, 170v, 206 and 207v, signatures on ff. 47r, 59r, 97v, 112r, 116v, 132r, 134v, 135r, 138v, 149r, 152v, 156r, 160r, 177r, 180v and 189v, and “a number of simple multiplication problems, calculations, annotations and scribbles [...] in the margins throughout” (Matheson 1999: 10.), all in the 17<sup>th</sup>-century hand identified with Marmaduke Towlard on the basis of the signatures.

In addition to these longer additions, there are also various shorter marginal scribbles, notes and pen-trials in the manuscript, apparently added by the different owners or readers of the manuscript in different periods. The following signatures are found in the manuscript in addition to the ones mentioned above. All of them are listed and dated in the BL MSS Catalogue or in Matheson (1999: 9–10):

*Rouland Brugge* (or *Bruges*), late 15<sup>th</sup> century (possibly the scribe), ff. 139r and 211r;  
*Henry Smythe of Gerlthorpe* (Garthorpe, Lincolnshire), late 15<sup>th</sup> or early 16<sup>th</sup> century, f. 211r;  
*Henry Smythe and Wyllam Lyddall* (both in the same hand), late 15<sup>th</sup> or early 16<sup>th</sup> century, f. 22v;  
*Thomas Roberteson*, early 16<sup>th</sup> century, f. 68v;  
*John Hall have deluud vnto Thomas Hall*, 16<sup>th</sup>–17<sup>th</sup> century, f. 106v;  
*John Towlard* (most likely written in the hand of Marmaduke Towlard), 17<sup>th</sup> century, f. 59r;  
*William Towlard*, 17<sup>th</sup> century, f. 158r;  
*John Thornton*, 17<sup>th</sup> century, f. 96v;  
*Richard Atkinson*, 17<sup>th</sup> century, f. 97v;  
*William Stowcroft*, 17<sup>th</sup> century, f. 158r; and  
*Ann Lee*, 17<sup>th</sup> century, f. 201v.

Small asterisks and large dots have been added next to some recipe titles (or the first line of the recipe) in a very black ink which is not that of the original scribe and could be either medieval or modern. The purpose for these markings is unknown. There are also some small marginal superscript letters *a* in the bills of fare (MS item 3b) next to some items, but these seem to be made in the hand of the original scribe.

### Origins and provenance

In terms of its dating, the manuscript contains several items and references that set a *terminus a quo* for the manuscript. First of all, it contains several items trans-

lated by John Shirley (b. *circa* 1366, d. 1456), an amateur scribe, book collector, and esquire (c. 1410-) and secretary (c. 1420-) to the Earl of Warwick, so it must post-date the beginning of his active translation career, from c. 1430 onwards. Based on these items, Christianson (1990: 234) has listed this MS as one personally associated with Shirley, but despite the similarity of their content, material, layout and decoration, to his personal anthologies, described by Boffey and Thompson (1989: 286), there is some internal evidence that would seem to suggest that the manuscript itself postdates his death. First of all, a reference to John Lydgate as a “late monke of Bury” on f. 221r means that it was written after his death in 1449 (Matheson 1999: 10), while a reference to Henry VI on f. 72v means that it was written after 1461. Based on the different paper used for MS items 1 and 2, the statement of the BL MSS Catalogue of these items being in a different hand than the rest of the manuscript, and the fact that MS item 3 begins on a new quire, it could be deduced that the manuscript was produced in two distinct parts (henceforth labelled A and B). Whether these two parts were initially or only later bound together is impossible to determine, but it is not implausible that MS items 3–10 were commissioned (or more unlikely, produced for sale) as a unit, while MS items 1 and 2 were then at some point bound together with this collection due to their similar instructional subject matter.

While this manuscript contains a large number of names written in its margins, it should be kept in mind that they do not in themselves constitute proof of ownership, but offer only circumstantial evidence at best (Harris 1989: 169–70). The BL MSS Catalogue conjectures that the name *Rouland Brugge*, written on f. 139r in a late 15<sup>th</sup>-century hand, and again on f. 211r in the form *Rouland Bruges*, could belong to the scribe who wrote part B of the manuscript. According to the manuscript catalogue of Ralph Thoresby, the manuscript “came from Selby, and is said to have belonged to the Monastery there”, although it is unclear at what point of its history. The name *Henry Smythe of Gerlthorpe* in an early 16<sup>th</sup>-century hand has been written on f. 208v, and the names Henry Smythe and *Wyllam Lyddall* on f. 22v. As the BL MSS Catalogue points out, Selby Abbey owned property at Garthorpe, Lincolnshire, which makes it more likely that this name refers to an early 16<sup>th</sup>-century owner—or at least reader—of the manuscript. The name *Thomas Roberteson* has been written on f. 68v in an early 16<sup>th</sup>-century hand, but it is uncertain whether it can be taken to indicate ownership or merely readership. On the other hand, the signature *H. P.*, dated 24 June 1597, which occurs on f. 149v has been interpreted as a sign of ownership by the BL MSS Catalogue. Unfortunately, the full name and identity of the signee are unknown. It is uncertain whether the inscription “John Hall have deluud vnto Thomas Hall” that occurs on f. 196r in a 16<sup>th</sup>- or 17<sup>th</sup>-century hand (BL MSS Catalogue) can be taken to indicate ownership or merely possession through e.g. lending. The note “Edward Parker booke” that occurs on f. 196r in a late 16<sup>th</sup>- or early 17<sup>th</sup>-century hand (BL MSS Catalogue), on the other hand would definitely seem to be a mark of ownership, although no date accompanies this signature, making it difficult to estimate the exact period of ownership. Four names, *John Thornton* (f. 96v), *Richard Atkinson* (f. 97v), *William Stowcroft* (f. 158r) and *Ann Lee* (f. 201v) have been inserted in the manuscript, each in a different 17<sup>th</sup>-century hand. It is uncertain whether these names indicate successive owners or merely readers of the manuscript over the 17<sup>th</sup> century, but they would seem to indicate that it was actively being read during this period.

However, the phrase “Marmaduke Towlard liber eius 1683”, written on ff. 112r, 132r and 135r (the last without the year) along with a number of other annotations in the same hand, most likely identifies the owner of the manuscript in this year. However, it is not known whether the date refers to the year of acquisition or merely the year in which he for some reason used and annotated the book.

At the end of the 17<sup>th</sup> century we have external evidence of the history of the manuscript, which was given to Ralph Thoresby by Thomas Wilson, a merchant from Leeds, before the year 1697, when it was listed as a part of Thoresby’s library by Bernard (1697: 230) (Thoresby 1715: 530). The manuscript was then sold at the auction of Ralph Thoresby’s museum on 7 March 1764 (lot 16) (Bristow 1764: 15), possibly to John Jackson living on Clement’s Lane, Lombard Street in London, to whom the manuscript belonged in the August of 1790 (Pinkerton 1797: 462). From Jackson’s collection, the manuscript ended up to the British Museum that purchased it on 30 April 1794 at the sale of Jackson’s manuscripts at Leigh & Sotheby, 28 April 1794, lot 340 (BL MSS Catalogue).

## 9.2.4 Oxford, Bodleian Library MS Ashmole 1439

### Physical description

The manuscript is written on parchment of varying colour, thickness and texture, ranging from near-white to grey and dark cream-coloured and from very smooth to heavily follicled. This might indicate that the parchment did not come from a single production batch but that the manuscript was written on miscellaneous leftover sheets. The folia have not been trimmed and vary somewhat in size, giving a ragged appearance to the edges of the manuscript. The original medieval portion of the manuscript consists of 55 small quarto (or large octavo) parchment folia to which a single initial leaf (f. 0) would seem to have been added later. The size of the leaves varies from 200 to 220 mm in height and from 130 to 150 mm in width. Ff. 41 and 42 are slightly narrower than the rest—only 130 mm—since they are single leaves (whose counterparts have been lost) that have been rebound by turning c. 10 mm of their inside margin into a stub and sewing through it. The thickness of the manuscript, excluding the covers, is 17 mm. In its current binding, the original parchment quires are both preceded and followed by nine flyleaves of fine white paper (and a tenth paper leaf has been pasted to the cover).

While the manuscript is generally in good condition, there is slight moisture damage evident at the edges of many pages, possibly indicating that the manuscript was at some point stored in slightly too moist conditions or was briefly subjected to water. The damage never penetrates more than c. 10 mm into the page, however. In addition to this systematic damage, there are also some occasional stains or droppings of unknown origin, which might testify to the use of the book outside a clean library setting. Between ff. 33 and 34 there is some yellowish staining near the top of the spine, extending c. 50 mm from the top of the page and c. 10 mm from the spine onto both folia. On ff. 33v there are also two brown stains with white residue, and a brownish stain at the bottom of the page. In addition to this light damage to the existing pages, there are several entire folia missing from the manuscript.

### Collation of the manuscript

The collation of the manuscript—excluding the modern paper flyleaves which each form additional quires of 5 bifolia—is as follows:

- 1<sup>6</sup>, single leaf (f. 0) added later (ff. 0-6);
- 2<sup>8</sup> (ff. 7-14);
- 3<sup>8</sup> (ff. 15-22);
- 4<sup>8</sup> (ff. 23-30);
- 5<sup>8</sup> (ff. 31-38);
- 6<sup>8</sup>, wants 3 and 4 (ff. 39-44);
- 7<sup>8</sup>, wants 4-6, two single leaves (ff. 50-51) added (ff. 45-51);
- 8<sup>0</sup>, four single leaves (?) added (ff. 52-55).

The rebinding of the manuscript at some point after the loss of some folia has introduced irregularities into its composition. While all of the pages seem to be in the right order, the sewing of quire 6 and the location of the missing folia do not coincide: while the quire is sewn between ff. 42 and 43, the folia are missing from between ff. 40 and 41. This means that at the time of rebinding, the two single leaves (ff. 41 and 42) which were originally the right-hand sides of their respective bifolia were ‘flipped around’ to the left and sewn from their verso sides instead of their rectos, the stubs being left to the right side of the middle of the gathering. The reason for the addition of the six empty single leaves (ff. 50-55) to the end of the manuscript is unknown. The first two of them are separately sewn onto ff. 48 and 49 using a stabbing stitch near the middle, but the method used for attaching the last four could not be discerned from the current binding.

**Signatures and catchwords** The first quire, which contains the incomplete table of contents and a number of empty pages, seems to have been produced separately from the rest of the recipe collection. It has no signatures, while the second quire seems to have been signed with *a*, followed by the numerals *j-ij* on its first four leaves, although they are so far at the bottom right hand edge of the recto sides that they are very difficult to discern. The third quire, on the other hand, clearly contains the signatures *b.j*, *b.ij*, *b.ijj* and *b.iiij* on ff. 15r-18r. For the fourth quire, the signature *c.j* on f. 23r is the only one that can be clearly discerned, although ff. 24-26 also seem to have signatures beginning with *c*. Quire 5 again preserves a complete series of signatures from *d.j* to *d.iiij* on ff. 31-34, while quires 6 and 7, as well as the empty single leaves seem to be entirely unsigned. Catchwords or catchphrases, written in the original scribal hand and replicating the first word or words of the next page, occur at the ends of quires 2, 3, 5 and 6 in the bottom margins of folios 14v (“a while”), 22v (“tempre vp”), 38v (“to gedre”), 44v (“chones”).

**Foliation** The manuscript contains two series of foliations, an earlier one in an italic hand writing in black ink (hand 6) at the right edge of the top margin, and a later one inserted in pencil to the bottom right corner of each recto. The reason for this double foliation is most likely the fact that the earlier foliation is defective in the sense that it runs from 1 to 34 and then jumps to 25, running all the way to 40 on f. 50r, while the later foliation runs continuously from 1 on f. 0r to 56 on f. 55r, foliating also the empty leaves at the beginning and end. The later foliator has also crossed out the earlier foliation in pencil. Since the earlier foliation also runs

uninterrupted past the missing leaves between ff. 40 and 41 and between ff. 47 and 48 both foliations were clearly inserted after the pages in question were lost.

### Manuscript layout

The layout is the same for the majority of the manuscript folia, consisting of a single column text block outlined and lineated in hairlines of brown ink; most likely the same as used for the quire signatures. The writing block measures 145–150 mm in height and 100 mm in width. The remaining space is divided between the margins, the top margin being c. 20 mm, the bottom margin measuring c. 45 mm and the inner and outer margins measuring 15 mm and 25 mm, respectively. The pages have been pricked with a sharp, round-profiled instrument—possibly a needle or an awl—and the pricks are still clearly visible at the very outer edges of the pages. The vertical lines outlining the edges of the text block reach all the way to the edge of the page, while the horizontal lines stop at the vertical lines or extend at most a few millimeters past them, except for the top line which occasionally extends to the edges of the page. The first 6 folia (of which only two have text on them) and the last written folio (f. 48) have been divided into two columns by a vertical line down the middle of the text block, and the table of contents and bills of fare contained on them have been written in two columns.<sup>19</sup>

The text has been written below the top line but decorative ascenders frequently used on the first line extend far into the top margin. The right edge of the text block is observed in principle, but in practice words frequently extend to the margin, the scribe seeming very reluctant to break words at line-end if there are only two or three letters that do not fit within the block. The pages have been fully lineated, 25–27 lines to a page, most pages having 26 lines with a line height of 5–6 mm. The baseline of writing is about 1 mm above the ruled line and the minimum height of letters is between 2 and 3 mm. Ascenders do not generally overlap with the descenders of the previous line, both terminating at a point 1–2 mm below the ruled line. In general, the visual appearance of the page is very well-spaced and ‘open’, which effect is enhanced by the very regular hand used for the text. As an exception, f. 0 has not been ruled at all, supporting the hypothesis that it is a later addition. The folia in the first quire containing the table of contents are written in two columns which are outlined by two additional vertical lines down the middle of the page, with a space of 12 mm in between. At the end of the manuscript, ff. 47v–48v have a single line down the middle of the page as a guide for writing the bills of fare in two columns.

In terms of marking textual organization, the recipes are written out in continuous prose with no line breaks, their rubricated titles serving to divide each recipe from the previous one.<sup>20</sup> In the table of contents, line fillers are used to fill up the space between each recipe title and its number, located at the right edge of the column.

<sup>19</sup> F. 47v, which contains the end of the last recipe and the beginning of the bills of fare, has not been ruled, but the bills of fare nevertheless been written in two columns.

<sup>20</sup> This type of textual layout, where the text runs as a continuous flow of text and individual recipes are marked only by rubricating their titles, is also exemplified by an extraneous quire added to California Huntington Library MS 1336, for which a digital image of a sample opening is available at <<http://digitalassets.lib.berkeley.edu/ds/huntington/images/001489A.jpg>>.

### Current binding

The current binding of the manuscript is of brown leather and shows some scuffing. On the spine, Elias Ashmole's coat of arms ("Quarterly sable and or, in the first quarter a fleur-de-lis of the second") within a laurel wreath is printed in gold, together with the shelfmark "ASH: | 1439." below it. The front and back covers, 222 mm by 152 mm in size, are plain. The thickness of the whole book, including the covers, is 25 mm.

### Summary contents of the manuscript<sup>21</sup>

- 1) An incomplete four-part numbered list of recipe titles, containing only the first two parts (ff. 1r-2v)
  - a) First part of list of recipe titles, containing 150 titles (ff. 1r-2r)
  - b) Incomplete second part of list of recipe titles, containing 52 titles (ff. 2r-2v)
- 2) A four-part collection of 269 culinary recipes (ff. 7r-47v)
  - a) A sub-collection of 152 recipes, titled *Potage* (ff. 7r-30r)
  - b) A second sub-collection of 64 recipes, titled *Leche viaunde* (ff. 30r-40r)
  - c) A third sub-collection of 34 recipes, titled *Viaundes ffurnes* (ff. 40r-45v)
  - d) A fourth sub-collection of 19 recipes, titled *Sauces pur diuerse viaundes* (ff. 46r-47v)
- 3) An incomplete collection of five menus (ff. 47v-48v)
  - a) A partial menu of three courses served at the coronation of Henry IV (f. 46r)
  - b) A partial menu of three courses, preserving only the third course (f. 48r)
  - c) A menu of two courses served for the lesser quests in lieu of the previous menu (ff. 48r)
  - d) A menu of three courses (f. 48r-v)
  - e) A menu of two courses (ff. 48v)

### Hands used in the manuscript

The manuscript is written entirely in a single hand but has some marginal and interlineal corrections and additions made in what would seem to be several hands, as well as two sets of foliation and some notes by later readers, cataloguers or librarians.

**Scribal hand** The hand of the original scribe is a very clear, compact and regular *textualis semiquadrata* with a minim-height of 2-3 mm, written in an ink whose colour varies slightly from dark black to dark brownish grey. The letterforms are quite typical to the script, including the capitals, which are often decorative in nature. Letters on the first and last lines of the page are also frequently provided with decorative ascenders and descenders. The *o*, *e* and *c* are quite pointed at the top in this hand, and the *w* has a single loop as its rightmost stroke. The scribe does not seem to make a distinction between *z* and *3*. In the table of contents, the scribe uses a variety of different capital letterforms for the initial letters of recipe titles. Since different forms of the same capital often occur in consecutive recipe titles, this would seem to be a conscious choice.

<sup>21</sup> A more detailed description of the contents of this MS is contained in appendix F and in the TEI header of the base data file for the MS included in appendix A.



**Other hands** In addition to the scribal hand, the following hands can be identified in the manuscript:

- 1) A small cursive hand in brown ink, used to prepare the parchment for writing by ruling the page in very fine lines of brown ink and adding quire signatures to the bottom right corners of recto pages to keep the folia in order.
- 2) The hand of the rubricator in red pigment, which is so similar to the scribal hand that it seems possible and even likely that the scribe him- or herself did also the decoration after having written the text.
- 3) A small 15<sup>th</sup>-century secretary hand in brown ink and of varying formality that has made corrections and additions both interlineally and in the margins. The interlineal corrections and additions would seem to emulate the hand of the original scribe in the sense that they are written quite carefully and formally, bordering on *textura*, but still with characteristic secretary letterforms. The additions and corrections made in the margin, however, are less formal and more cursive. This could mean that there are actually two annotators at work here, but since the ink used for both of the styles looks exactly the same and the formality of the hand seems to fluctuate quite fluidly (with many examples whose formality lies somewhere between the two extremes), it seems more likely that both the interlineal and marginal comments are by the same person. One explanation would be a single person (perhaps even the original scribe) doing two rounds of revision and correction, using different pens (and thus hands) but the same ink.
- 4) A post-medieval italic hand in brownish black ink that has written the phrase “Henry Gournay me possidet” at the top of f. 0r.
- 5) A modern italic hand in black ink that has added recipe numbers in the margin for recipes 2-14 (on ff. 7r-9r), and most likely also the notes “A°1” on f. 0r, “circa 1400” on f. 0v and “A1439” on f. 49r. This hand could conceivably belong to Elias Ashmole or someone responsible for his library. This hand would seem to be different from the one that has added the foliation (hand 6).
- 6) A post-medieval italic hand in black ink that has added the defective foliation to the top right hand corner of the recto pages. This hand would seem to be different from the one that has added the recipe numbers and cataloguing notes (hand 5).
- 7) A modern hand in pencil that has corrected the foliation of the manuscript by crossing out the old foliations in ink (hand 6) at the top right corners of the folio pages and writing new ones in pencil to the bottom right corners of the pages. This hand would seem to be different from the one that has added the catalogue information on the inside front cover (hand 8).
- 8) A modern hand in pencil that has added the catalogue entry “SC 8342” to the top left corner of the inside front cover and the MS signum “MS. Ashmole 1439” to the center of the same page.
- 9) A decorative italic hand in brown ink that has added the word “Potage” centered above the first ruled line of the writing block on the otherwise empty f. 6r. The colour of the ink is very similar to that used for ruling the pages, but the hand is decidedly modern.
- 10) An early modern secretary hand in yellowed ink that has made some pen trials to the top margin of f. 43r.
- 11) An unknown hand that has made a single erasure and could be the same as any other hand involved with the manuscript.

## Decoration

In line with its parchment material, regular layout and neat hand, this manuscript is also relatively decorated, mainly through the use of rubrication, but also of decorative ascenders and descenders frequently used on the top and bottom lines of the page. The main decorative feature of the text is the use of red pigment both

for highlighting textual features and seemingly purely for ornamentation. Titles, both for the different recipe groups and for the recipes themselves are written in red, as are the recipe numbers found in the table of contents. Furthermore, the initial letters of recipes are frequently highlighted by a stroke of red pigment. The ornamental use of red pigment takes the form of line fillers constructed of curving oblique strokes joined together into a braid-like pattern. These are used mainly in the table of contents to fill the space between the recipe title and its number, but occasionally also between recipes or between a recipe title and the recipe itself, where they are often combined with various punctuation symbols. It can be clearly seen that the rubrication was added after the text itself had been written, since the original scribe seems to have used faint cross-like symbols in the margin to indicate the places to be rubricated. Furthermore, the space left for the rubricated titles is frequently too short, resulting in the titles overlapping either with the final punctuation of the preceding recipe or the first letter of the current recipe.

While no proper decorative initials are used, the initial letters of recipe titles—themselves written in a slightly more formal hand than the recipes themselves—are often quite carefully executed Gothic capitals with several purely ornamental strokes. In addition to these initials, many letters on the top and bottom lines have decorative ascenders or descenders added to them. This tendency for the elaboration of letterforms also extends to punctuation; *puncti* frequently have a decorative tail stroke, which does not seem to imply any functional difference.

### Later additions

The manuscript does not contain extensive additions or annotations, but there are some marginal and interlineal corrections and additions that seem to postdate the initial writing of the manuscript, as well as two sets of modern foliation (discussed above) and some notes by later readers, cataloguers or librarians. The earliest layer of additions seems to be the interlineal and marginal emendations that have been done in a 15th-century secretary hand that at times clearly tries to emulate the style of the original scribe but uses different letterforms than the original scribal hand (hand 3). These emendations may be either a part of the production process of the manuscript, being added by a corrector, or they may have been added by an early user of the manuscript. The neat and professional character of the hand would suggest the former.

There are also several minor annotations made to the manuscript by later owners or readers of the manuscript. None of these bear any indication of their date, but based on their hands they have all been made after the medieval period. These include the word “Potage” added in a decorative italic hand writing in brown ink above the top line of the otherwise empty f. 6r and the pen trials made in the top margin of f. 43r in an early modern secretary hand in yellowed ink. The most tantalizing of these later markings is the owner’s inscription at the top of f. 0r, which states in an italic hand that “Henry Gournay me possidet”. In the absence of a date (Black (1845) notes this inscription but makes no judgement as to its dating) or some external evidence, the identification of the said Henry Gournay remains difficult if not impossible.

Other later annotations include the addition of marginal recipe numbers for

recipes 2-14 and some bibliographical notes on f. 0r ("A°1"), f. 0v ("circa 1400") and f. 49r ("A1439") in an italic hand in modern ink, which could conceivably have been added by Elias Ashmole himself or someone responsible for his library. The hand that has added the foliation, although also in black ink, would seem to be different from this hand. The catalogue number "SC 8342" and the signum "MS. Ashmole 1439" have been added onto the inside of the front cover in pencil, supposedly by a librarian of the Bodleian Library sometime during the 19<sup>th</sup> or 20<sup>th</sup> century, possibly in 1860 when Ashmole's manuscript collection was transferred from the Ashmolean Museum to the Bodleian Library.

### Origins and provenance

The only direct internal basis for dating the manuscript is the mention of Henry IV's coronation in 1399, but the fact that the bills of fare elsewhere connected with bishop John Stafford's consecration in 1425 (MS item 3b) and the wedding of the Earl of Devonshire c. 1430 (MS item 3d) are included in the manuscript means that it must postdate these events.<sup>22</sup> While Black (1845: 1087-8) dates the manuscript to "early in the XVth century" and Austin (1888: vii) remarks that this manuscript "is about the same date as Harleian MS. 279", i.e. 1430-1440, the exhibition catalogue of Elias Ashmole's tercentenary exhibition (Hunter et al. 1983: 74) revises this dating to a more conservative estimate and dates the manuscript to the second half of the fifteenth century, which is a reasonable estimate taking into account the aforementioned internal evidence.

At some point in its history, most likely in the 16<sup>th</sup> century and certainly before Elias Ashmole's (1617-1692) acquisition of it, the manuscript was apparently owned by a Henry Gournay, who wrote his name on f. 0r of the manuscript. The manuscript was acquired by Elias Ashmole at some point during his active career as a collector of curiosities, books and manuscripts, most likely after his marriage to lady Mainwaring in 1647, which provided him with the means to pursue his interests as a collector of books and curiosities, and certainly before his death in 1692 (Madan, Craster and Young 1937: 1115; Hunter et al. 1983: 3). From 1683, when the building of the Ashmolean Museum was completed, to 1860 this manuscript was housed in the Ashmolean Museum, after which it was moved to the Bodleian Library, along with the rest of the Ashmolean manuscript collection.

## 9.2.5 Oxford, Bodleian Library MS Douce 55

### Physical description

The manuscript is a small volume written on thin parchment varying in colour from white through a dark creamy colour to brown. In addition to the colour, also the thickness and stiffness of the parchment varies between different sheets, as well as its translucency. On many pages the parchment is so translucent that the writing on the other side of the sheet is clearly visible, making the text difficult to read. It is unclear whether this translucency is due to mistakes in the prepa-

<sup>22</sup> Based on the observation that some of the information regarding the context of the bills of fare has been left out, it could be conjectured that this version of them is later than the one found in MS H279, supposing that the information is accurate and not a later fabrication.

ration process of the parchment or to subsequent conditions having affected it, but according to Hector (1966) this kind of “an appearance of greasiness or semi-transparency” (16) is characteristic of poor quality parchment. The original part of the manuscript consists of 80 small octavo (or large 16mo) parchment folia, c. 135 mm high and c. 95 mm wide. The pages of the manuscript do not seem to have been trimmed after being bound to the current binding, but they have been trimmed at some earlier stage. The modern flyleaves preceding and following the original manuscript in the current binding are of paper with a large watermark of a large creature holding a sword and surrounded by a wooden palisade. A large number of the leaves have been mutilated at some point during its history, ff. 1 and 2 missing their outer halves and ff. 3-33 missing their bottom quarter.

The parchment used for the manuscript is not of very high quality and has suffered various kinds of wear and damage during its history. The stiffness, brittleness and translucency of the parchment has resulted in wear on the text. This has been exacerbated by a very tight binding, which has caused the folia to cockle and rub against each other. The resulting wear is especially visible in the initials painted in blue pigment, which have in many places been rubbed off to a large extent. For some reason ff. 49-56 are even more translucent and brittle than the rest of the manuscript and thus quite difficult to read. There is also quite significant staining visible in the manuscript, both in the form of clear circular stains that seem to be the result of something dropping on the page and more general smearing. The most obvious form of damage suffered by the manuscript is the excision of the outer halves (c. 55 mm) of ff. 1-2 and the bottom quarters (c. 40 mm) of ff. 3-33, which has resulted in the loss of the beginnings or ends of lines for much of the table of contents, and of one to three lines of text at the bottom of each page. This excision is commented upon by Francis Douce in an inscription on f. iir, where he attributes the damage to a “keeper of a chandlers shop” in Chertsey, who had used it as a source of parchment for making labels; presumably intending to use the empty bottom margins of the pages but carelessly cutting off a few line of text as well. Apart from these mutilated pages, the manuscript seems to have been preserved in its original collation with no loss of entire pages, although one of the final flyleaves (f. 83) has come loose from the binding.

### Collation of the manuscript

In the current binding, the original manuscript quires are preceded by 6 (i8 wanting 1 and 2) and followed by 7 (ii8 wanting 8) modern paper flyleaves. The quiring of the original parchment leaves is difficult to establish with certainty, as the tight binding hides the threads used for stitching it together. The collation of the manuscript, excluding the modern flyleaves, is as follows:<sup>23</sup>

- 1<sup>8</sup> (ff. 1-8);
- 2<sup>8</sup> (ff. 9-16);
- 3<sup>8</sup> (ff. 17-24, f. 17 foliated as 16\*, following MS foliation thus behind by 1);
- 4<sup>8</sup> (ff. 25-32, foliated 24-31);
- 5<sup>8</sup> (ff. 33-40, foliated 32-39);
- 6<sup>8</sup> (ff. 41-48, foliated as 40-47);

<sup>23</sup> The folio numbers reflect the actual manuscript structure and not the modern foliation found in the manuscript.

- 7<sup>8</sup>, slip of paper (f. 53) added (ff. 49-57), foliated as 48-55;
- 8<sup>8</sup> (ff. 58-65), foliated as 56-63;
- 9<sup>8</sup> (ff. 66-73), foliated 64-71;
- 10<sup>8</sup>, 1 single leaf (f. 82) added (ff. 74-82), foliated as 72-80.

**Signatures and catchwords** Any signatures there may have been on the first four quires have been lost with the damage to the outer and lower parts of the folia, but there is an almost complete series of signatures on quire 5, ff. 34r, 35r, and 36r carrying the signatures *e2*, *e3* and *e4*. It is likely that also the first four quires were signed, as ff. 42 and 44 have signatures beginning with *f*, f. 44 preserving also the numeral 4 following it, and f. 52 has a signature beginning with *g*, implying a consistent signature scheme for the entire manuscript. There are two surviving catchphrases in the manuscript, on ff. 40v (“kydde or henne”) and 73v (“hem yn|water”), of which the latter contains also the last word of f. 73v in addition to the first two of the next page.

**Foliation** The manuscript has been foliated with arabic numerals at the top right hand corner of each recto page in an italic hand that resembles that used for the inscription by Francis Douce. As the foliation has been added after the loss of the outer halves of ff. 1-2. It would thus seem likely that Douce himself foliated the manuscript when he had acquired it in 1787. The foliation is defective in that the number 16 is repeated twice, the second instance being distinguished with an asterisk, added most likely after the foliator noticed the mistake. The slip of paper added between ff. 52 and 54 (51 and 52 in the MS foliation) has been foliated as “51b”. These anomalies mean that the manuscript foliation diverges from the foliation used in this edition, which follows the actual structure of the manuscript.

### Manuscript layout

The layout is the same for the majority of the manuscript folia, consisting of a single column text block—outlined in light brown ink with lines reaching to the edges of the page—measuring 80-90 mm in height and c. 60 mm in width. The remaining space is divided between the margins, the top margin measuring c. 16 mm, the bottom margin measuring c. 30 mm and the inner and outer margins measuring c. 6 mm and c. 25 mm, respectively. The ruling is very faint, and on many pages all that is visible is top line and short sections of the vertical rules around it. There is occasional evidence of pricking for the top line in the outer margin, although the pricks have in many cases been trimmed away. The pricking would seem to have been done with a thin knife or some other flat implement, leaving a narrow horizontal slit.

Although the page has not been lineated for individual lines of text, the lines of text are relatively horizontal and regular in spacing. The number of lines per page varies, but the average is around 20 lines of text per page, making the average height of the writing line c. 4 mm. The average minim height of the scribal hand is around 1.5 mm, making for a somewhat crowded appearance, the ascenders and descenders overlapping significantly. The first line of text has been written below the ruled top line, and also the bottom rule is quite well observed, there frequently being almost a full line height of space between the bottom of the last line and the

ruled line. The right-hand side rule is observed less strictly, a significant proportion of lines crossing it and extending into the margin. On ff. 1r-5r, which hold the table of contents, two additional vertical rulings have been added, dividing the writing block into three columns of c. 8mm, 14mm and 40 mm in width, the first for the number, the second for empty space and the third for the recipe title. As the recipes are numbered in roman numerals, the column for the recipe number is frequently too narrow to hold the number.

In terms of marking textual organization, individual recipes are separated from each other by empty space, the height of which varies from a single line to the equivalent of three lines of text. Final lines of recipes that do not fill the entire line are not filled but remain empty. The recipe titles are added by the rubricator more or less centered in the space left between the recipes, while recipe numbers are located in the outer margin next to the first 3 lines of the recipe itself.

### Current binding

The manuscript is bound in brown leather decorated with gilt tooled lines, and was most likely bound into its current binding already upon its acquisition by Douce, as his *Ex Libris* is affixed to the inside of the front cover. The spine of the book is very worn, with the brown fabric used to build it visible in many places beneath the leather. There are some remains of gilt lettering on the spine, of which the letters “IE || ERY” are visible.

### Summary contents of the manuscript<sup>24</sup>

- 1) A damaged numbered list of originally 185 recipe titles with subheadings (ff. 1r-5r)
- 2) A five-part collection of 184 culinary recipes with a main part containing 98 recipes and four titled subgroups (ff. 5r-81r)
  - a) A sub-collection of 14 recipes, titled *Diuerses sauces pur diuerses viaundes* (ff. 45v-50r)
  - b) A sub-collection of 24 recipes, titled *La manere pur roster et saucer diuerses viandes* (ff. 50r-57v)
  - c) A sub-collection of 21 recipes, titled *La maner pur roster fryer et builler diuerses pessones* (ff. 57v-64r)
  - d) A sub-collection of 34 recipes, titled *Diuerses sauce en quaresme* (ff. 64r-81r)

### Hands used in the manuscript

The manuscript is written entirely in a single hand with decoration and highlighting in colour, possibly by the same scribe. There do not seem to be any contemporary emendations or annotations but a modern foliation has been added, as well as some notes by later readers, cataloguers or librarians.

**Scribal hand** The hand of the original scribe is a reasonably neat but compact and rather squat, rounded and somewhat cursive secretary-anglicana hybrid written with a relatively wide nib and greyish black ink, which in some places has turned brown, possibly due to water damage. The ink seems to have aged very unevenly,

<sup>24</sup> A more detailed description of the contents of this MS is contained in appendix F and in the TEI header of the base data file for the MS included in appendix A.

as the colour of the ink varies quite significantly, being cold grey in some parts of a page and warm brown in others.<sup>25</sup> The minim-height of letters is c. 1.5 mm, line height being around 4 mm. In terms of the letterforms used, the hand is a true hybrid between secretary and anglicana. It uses exclusively the pointed single-compartment secretary *a* and the 'traditional' lobed minuscule *e* typical to secretary, as well as the secretary *g* with the open descender and the horizontal top stroke. From anglicana, it adopts the sigma-shaped word-final *s*, the *d* with the looped ascender and the relatively complicated anglicana *w*. The hybridity is also evident on the level of some individual letters, like the *r*, of which the hand uses both the short secretary variant and the long version typical to anglicana (both forms exhibit a varying degree of lateral 'openness'). The letter *y* is marked with a round dot above it.

**Other hands** In addition to the scribal hand, the following hands can be identified in the manuscript:

- 1) The hand of the rubricator, who has added recipe numbers (both in the table of contents and the collection itself), recipe titles (in the collection itself), paragraphs and underlining in red pigment, and decorative large initials in blue pigment. The hand of the rubricator is very similar to that of the original scribe and it is likely that the decoration was done either by the same person or a colleague with a very similar scribal style. The fact that the decoration was added afterwards is revealed by the small minuscule letters written by the original scribe as instructions for the rubricator, still visible beneath some of the decorative initials.
- 2) An italic hand in brown ink, possibly belonging to Francis Douce himself, that has foliated the leaves in arabic numerals at the top right hand corner after the loss of the outer halves of ff. 1-2. Both the ink and the shape of numbers used for this foliation are very similar to those found in the inscription by Douce on f. iir, and it is not unlikely that he would have foliated the manuscript upon acquiring it in 1787.
- 3) A small, upright and quite rigid modern hand in pencil that made some markings in the margins, including the folio number "51 b (ult)" on the note slip added between ff. 52 and 53 and the folio numbers 81-87 on the empty ff. 83-89. The hand most likely belongs to a librarian or other custodian of the manuscript.
- 4) An italic hand in pencil that has added what appears to be a list of folio references to f. 88v. The significance of this list of folio numbers is unclear.
- 5) An early 19<sup>th</sup>-century italic hand in black ink that has written the note slip added between ff. 52 and 53.
- 6) An early 18<sup>th</sup>-century italic hand in black ink, apparently belonging to J. Urry, which has added a lending note on f. 82v, stating that the writer borrowed the MS from a *Mr J. Chicheley* on the 24<sup>th</sup> of May 1714.

## Decoration

Despite its small size and low-grade parchment, this manuscript contains several decorative features, mostly involving the use of colour, that also serve a text-organizing function. The most prominent of these are the two-line Uncial-style initials in blue pigment that are used to mark the beginning of each recipe. Some of these initials (on ff. 5r, 5v and 8r) are further decorated with red hairline scrollwork surrounding them. In addition to the initials, the table of contents begins with an equally large blue paragraph (f. 1r) similar in style to the initials and also quite elabo-

<sup>25</sup> Although this might also be due to the varying quality of the parchment, which has absorbed the ink in different ways in different places.

rately decorated with red scrollwork. Apparently the blue initials were added by a separate rubricator after the text had been written, since the small minuscule letter in ink, intended as an instruction for the rubricator, is still visible underneath or next to the blue initial. In some cases the rubricator has missed the instruction and the original guide letter can be seen in the middle of an empty space reserved for the initial.

In addition to the blue initials, red pigment has also been used for emphasis. All recipe numbers, both in the table of contents and the collection itself, are written in red pigment, as are the recipe titles themselves in the recipe collection (in ink in the table of contents). In both the table of contents and the collection, most (but not all) recipe titles are preceded by a red paraph, and in the recipe collection itself most of both titles and recipe numbers are also underlined in red pigment. In the margins, where recipe numbers have been added by the rubricator, are traces of the recipe number having originally been written in a small script in brown ink and subsequently erased by the rubricator. On ff. 61v and 62r these original markings have survived, there being neither rubricated recipe numbers nor recipe titles, most likely due to the rubricator having accidentally turned two pages at a time. The fact that the blue initials have been added reveals that the decoration was done in two passes, most likely first adding the initials in blue and then adding the numbers and titles in red.

### Later additions

In what would be the top margin of f. 82v (i.e. the last page of the original manuscript) there is an 18<sup>th</sup>-century inscription in an italic hand: “24th May 1714 borrowed this || of *mister John chicheley* of the midle Temple || by me *John Urry*. 1714”. The ex libris of Francis Douce, consisting of the arms of Francis Douce—“Or, a chevron chequy azure and argent between three greyhounds courant sable, a mullet for difference”—has been pasted to the inside of the front cover. There are also several inscriptions on the modern flyleaves preceding and following the original manuscript, including the shelfmark “Douce || ms. || 55” in the middle of the inside front cover and the catalogue number “S.C. 21629” in the top left corner of the same page. On f. iir (the second modern flyleaf), there is an inscription in an italic hand presumed to belong to Douce himself:

For the possession of this curious ms I am indebted  
to the kindness of M<sup>r</sup>. Smith apothecary at Chertsey,  
who rescued it from the hands of the keeper of a  
chandlers shop in that town who had begun  
to cut it in shreds to make labels

J. Douce 1787.

A small (c. 53 mm wide and 92 mm tall) slip of paper, torn from an unidentified document, has been inserted between ff. 52 and 54 (and foliated in the present edition as f. 53). On its recto side is following inscription in an italic hand and black ink: “Inscription on a large culinary vessel exhibited to *sosiety* of *Antiquaries* May || 1801 by *Colonel Greville*. || IE SVS POT DE GRANT HONVR || VIAVNDE A FAIRE DE BON SAVVR .”. On its verso side are remnants of undecipherable vertical lines of text and a foliation “51b || (ult)” in pencil. On f. 88v (the second



to last modern flyleaf), an italic hand in pencil has added (hand 4) a vertical list of what would seem to be folio numbers, consisting of the following items: 7 *b*, 10 *b*, 16 *r*, 17, 19, 22 *rb*, 27 *b*, 31 *b*, 32, 47, 48, 55, 65 *b*, 69 *b* and 78. The purpose or meaning of the list is unknown.

### Origins and provenance

There is no internal evidence in the manuscript that would allow it to be dated. Madan, Craster and Young (1937: 506) and Gillam (1984: 140) date the manuscript to the mid-fifteenth century, which is quite plausible in terms of both the content and the scribal hand of the manuscript. Based on the inscription on f. 82v, the manuscript seems to have been acquired by “mr J chicheley” sometime before 1714. This is most likely John Chicheley, “son of rear-admiral Sir John Chicheley (d. 1691), [who] both matriculated at Christ Church, Oxford, and was admitted to the Middle Temple, in March 1694/5, aged 16” and “was later a fellow of All Souls (res. 1703), and was called to the Bar in February 1701/2” (Gillam 1984: 140, see also Foster 1891: 270). In 1714, the manuscript was lent by its then owner, John Chicheley to “J. Urry”, who is most likely John Urry (d. 1715), the Chaucer editor (<<http://www.oxforddnb.com/view/article/28021>>). If the story recounted in Douce’s inscription is given credence, the manuscript had then somehow ended up in the hands of “keeper of a chandlers shop” in Chertsey sometime before Douce’s acquisition of the manuscript.

According to the same inscription, the manuscript was then rescued by a certain Mr. Smith from a chandler who had started to cut it into strips to use as labels, and delivered to Francis Douce in 1787, which would make it one of his earlier acquisitions. Judging from the inscription, it seems likely that the manuscript was given to him by the “M<sup>r</sup>. Smith apothecary at Chertsey”, rather than sold.<sup>26</sup> Upon Francis Douce’s death on March 30, 1834, his entire collection of manuscripts, including this one, was bequeathed to the Bodleian Library, along with his collections of printed books, coins and prints.

## 9.2.6 Durham, Durham University Library MS Cosin V.iii.11

### Physical description

The manuscript is a composite one, consisting of nine distinct *booklets* (see subsection 8.2.2), here referred to as parts A–I, each consisting of one or more quires. Several—if not all—of these booklets were collected together already in the 15<sup>th</sup> century, with some possibly in the 17<sup>th</sup> century when the manuscript was bound into its current format for George Davenport. The manuscript is composed of a total of 107 leaves of Chancery quarto size (Gumbert 2000: 81–2), measuring approximately 218 by 150 mm. Of the folios, the first (f. i) and the last (f. 105) are from the 19<sup>th</sup> century, while ff. ii, 26–29, 57–60 and 101–104 are from the 17<sup>th</sup> century. The rest, 92 folia, are original 15<sup>th</sup>-century leaves. Apart from the outer and inner bifolia of part E (ff. 61, 66–67 and 72) which are of parchment, the manuscript is written on several types of paper. The following watermarks have been

<sup>26</sup> However, Gillam (1984: 132) points out that the information in Douce’s inscriptions—some of which were added later—is not always entirely reliable.

identified by Doyle (1992) in the different parts of the manuscript:<sup>27</sup>

- A) twin: Colonne, with three steps at each end, cf. Briquet (1923) number 4347 and Mosin and Traljac (1957) number 3118, but both dated uses are from 1353; the former says the watermark is of an Italian type, ceasing 1364–1425, thereafter other variant types;
- B) also Colonne as above;
- C) Etoile, cf. Briquet (1923) numbers 6017–8, used in Flanders and France 1412–16;
- D) circle and cross (uncertain);
- E) three mounts with cross, cf. Briquet (1923) number 11678 et seq., Heawood (1930) Sources 67 (Durham 1435–56);
- F) ff. 75–76 and 82, cf. Basilic, Briquet (1923) numbers 2648–82 (1389–1512), Drache, Piccard (1961–1997) X.ii.386–669 (1393–1492), Heawood (1930) Sources 88 (1378, 1411, 1453); f. 77, three mounts with cross as in (E);
- G) none;
- H) none;
- I) a quadruped passant (horse with head turned back), not found in Briquet (1923) or Piccard (1961–1997);

Unfortunately, the usefulness of watermarks as evidence of the origin of the manuscript is limited by the fact that where we cannot identify the exact same *moulds* being used, the general *type* of a watermark can “at best only be broadly indicative” and at worst downright misleading (Lyll 1989b: 17–8, 1989a: 251).

Almost all leaves of the manuscript (including the 17<sup>th</sup>-c. ones but not the 19<sup>th</sup>-c. ones) have suffered moisture damage along the top and upper fore-edge. According to Doyle (1992: 1), this has probably occurred in Cosin’s library either during or after Davenport’s time (late 17<sup>th</sup> c.). Based on the similarity of the moisture stains throughout the manuscript, it probably occurred after the manuscript was bound in its current configuration. Judging from the multiple lines left by the moisture, it would also seem that the manuscript was exposed to moisture at least twice, having dried at least to some degree before getting moist again. There also seems to be some moisture damage near the middle of the spine edge in quires 1 and 2 (ff. 1–21). The date of this damage is unknown, but it could mean that the said quires were at some point stored without a binding, whose spine would probably have protected this portion of the quire from moisture. Also quire 3 (ff. 22–25) probably spent some time as a loose quire, either together with quires 1 and 2 or separately, since f. 25v is rather dirty along the spine edge, as if it had for a time been the outermost folio of an unbound sheaf. For an unknown reason, also ff. 16r and 17v (the outsides of the middle bifolium of quire 2) are similarly dirty along the spine edge.<sup>28</sup>

Several holes of various sizes (1–3 mm), most likely caused by bookworms, extend through sections A and B (ff. 1–25) but not through the 17<sup>th</sup> c. leaves

<sup>27</sup> The 17<sup>th</sup>-century leaves (ff. ii, 26–9, 57–60 and 101–4) have a watermark depicting a fleur-de-lys in crowned shield with maker’s monogram WR below, described in Heawood (1930).

<sup>28</sup> As Denholm-Young (1954: 60) points out, this is not a rare occurrence, as a completed work often had “a considerable existence in the form of loose quires before binding”, especially in the case of shorter works which would later be bound up into a larger volume together with other works, sometimes constituting groups or sequences of texts that could then become regularly copied as a whole (Edwards 1998b: 162–3).

(ff. 26–29). There does not seem to be bookworm damage in the other parts of the manuscript, indicating an early association of parts A and B before the manuscript was bound to its current configuration. This damage may have occurred during the same period as the moisture damage near the spine edge of A and the scuffing on f. 25v. Some of the bifolia of A and B are also damaged at the spine, around the binding, but whether the damage is due to tearing or bookworms is hard to say. Both the holes and the damage near the spine affect text only in rare cases, all of which are indicated in the transcription. In addition to the moisture damage and the wormholes, there is also some evidence of use in the form of staining—most likely from dirty fingers—especially at the lower half of the outer margin, which is the most natural place for the reader to grab when turning the pages.

### Collation of the manuscript

The collation of the manuscript, excluding the 19<sup>th</sup>- and 17<sup>th</sup>-century leaves and divided into its composite parts, is as follows:

A. ff. 1–21:

1<sup>14</sup>, wants 4 and 11 (ff. 1r-12v);

2<sup>8</sup>, 1 single leaf (f. 21) added (ff. 13r-21v).

B. ff. 22–25:

3<sup>6</sup>, wants 1 and 6 (ff. 22r-25v).

C. ff. 30–46:

1<sup>18</sup>, wants almost all of 1, fragment of it stuck between ff. 45 and 46, (ff. 30r-46v).

D. ff. 47–56:

5<sup>12</sup>, wants 9 and 10, stubs transposed and stuck to f. 51 (ff. 47r-56v).

E. ff. 61–72:

6<sup>12</sup> (ff. 61r-72v).

F. ff. 73, 75–82 (bound together with item G below):

7<sup>16</sup>, wants 5–6, 13–14 and 16 (ff. 73, 75–82).

G. ff. 74, 83:

7<sup>2</sup>, a single bifolium intruded into part F (ff. 74 and 83).

H. ff. 84–93:

8<sup>12</sup>, wants 1 and 12, ff. 88–89 of earlier origin (ff. 84r-93v).

I. ff. 94–100:

9<sup>6</sup>, 1 single blank leaf (f. 21) added (ff. 94r-100v).

**Signatures and catchwords** In MS part A, ff. 2r-6r have signatures *ija-vija* drawn with a fine pen in red, possibly by the foliator, while ff. 14r-16r have signatures *ij-iiij* drawn with a fine pen in light brown ink, which could well be the same used for the text of this part. In MS part D, ff. 47–52 have been signed by the scribe, *j.a* - *a.vj*. In addition ff. 47, 52 and 54–56 have also been signed by a different 15<sup>th</sup>-c. hand in darker ink, *jg*, *6g* and *8g-10g*, (i.e. after the loss of the two leaves after f. 54). In MS part F, ff. 73r, 75r, 76r, 77r and 78r have been signed, respectively, *aj*, *bij*, *cijj*, *Fvj* and *g*, possibly by the scribe and certainly before the insertion of f. 74 and f. 83 and the loss of two bifolia (between ff. 76 and 77 and ff. 82 and 83). These leaves have also been signed in the same darker ink as part D above (ff. 47–56), *j o*, *3 o*, *4 o*, *7 o*, *8 o*. This means that these second signatures were added after the insertion of f. 74 and f. 83 but before the loss of the two bifolia. F. 74, inserted into

MS part F, has also been signed 2 o in the same darker ink as in D and F above.

**Foliation** The only section of the manuscript that bears original 15<sup>th</sup>-century foliation is ff. 30–40, which have been foliated 1–11 in ink in the top right corner of the recto side of each folio. During the 18<sup>th</sup> century, ff. 1–21, 22–25 and 61–72 have been separately paginated in ink in the top right corner of the recto side of each folio by Robert Harrison. The entire manuscript has been foliated as i–ii and 1–105 in pencil in the top right corner of the recto side of each folio by I. A. Doyle c. 1951, which is also the foliation followed by this edition.

### Manuscript layout

Since the layout of the manuscript varies between the different parts, the layout of each part is described separately.

- A) Pages are framed in brown ink (possibly the same used for the text itself), the size of the writing block varying slightly, being on average 180 mm in height and 115 mm in width. The top margin varies from 11 to 14 mm, and the bottom margin from 28 to 38 mm, while the inner and outer margins are roughly 15 mm and 20 mm wide, respectively, the former curving quite strongly into the fold. The menus and the table of contents on f. 1 are laid out in three roughly equal columns, each having 41–45 lines, while the recipes themselves on ff. 2–21 are laid out in a single column having 31–41 lines of varying height per page, the average line height varying between 4 and 5 mm. The minim height of letters is between 1.5 and 2 mm, with moderate ascenders and descenders, the ascenders overlapping with the descenders of the previous line. In general, the visual appearance of the page is that of rather dense text blocks for individual recipes well-separated from each other. The lines have not been ruled and the written lines are not always perfectly horizontal. Neither is the right side of the frame always observed by the scribe, the text frequently extending beyond the frame. In terms of marking textual organization, the recipes are clearly separated from each other by an empty line of variable height, on which the titles are written, often extending partially to the margin outside the text block. Also the 18<sup>th</sup>-c. recipe numbering has been added on this empty line, roughly centered on it.
- B) Pages are framed in ink, the size of the writing block varying slightly, being on average 180 mm in height and 113 mm in width. The recipes are laid out in a single column of 40–44 lines per page. The recipes are clearly separated from each other by an empty line of variable height, on which the titles are written. Also the 18<sup>th</sup>-century recipe numbering has been added on this empty line, roughly centered on it.
- C) Pages are framed in brown ink, the size of the writing block varying slightly, being on average 180 mm in height 110 mm in width. Text is laid out in a single column of 34–39 lines per page.
- D) On ff. 47–54 there is no sign of framing or ruling on the pages, the size of the written space measuring c. 170 by 105 mm in width, divided between 3 columns of 31–32 lines per page. Ff. 55–56 are also unruled with a written

measuring 170–201 mm in height and c. 125 mm in width, with the text laid out in a single column of 24–29 lines per page.

- E) Pages are framed in softish brown ink, delimiting a writing block 173–185 mm in height and 120–125 mm in width, with the text laid out in a single column of 32–33 lines per page.
- F) There is no sign of framing or ruling on the page, the size of written space measuring 192–207 mm in height and 120–134 mm in width, with the text laid out in a single column of 42–51 lines per page.
- G) There is no sign of framing or ruling on the page, the size of the written space measuring up to 183 mm in height and 130 mm in width, with the text laid out in a single column of up to 35–37 lines per page.
- H) There is no sign of framing or ruling on the page, the size of written space varying widely, being between 150 and 200 mm in height and between 105 and 125 mm in width. The text is laid out in a single column with 23–25 lines per page.
- I) There are traces of framing in sharp grey ink on the page, the size of the writing block being 170 mm by 115 mm with text laid out in a single column, 26–28 lines per page.

### Current binding

The current binding of the manuscript is not contemporary with its contents but dates from the 17<sup>th</sup> century. George Davenport (c. 1631–1677)—a chaplain of John Cosin, Bishop of Durham (1595–1672)—acquired the manuscript for the Bishop's library and had it bound in its present binding of brown speckled calf, with blind double fillets and a roll on the sides. In the mid-19<sup>th</sup> century, the spine of the book was replaced and a brass clasp added to keep it closed.

### Summary contents of the manuscript<sup>29</sup>

#### A

- 1) Two menus and a list of recipes (ff. 1r-1v)
  - a) A menu of three courses served to Richard II (f. 1r, col. a-b)
  - b) A menu of three courses served to the king (f. 1r col. b)
  - c) A list of 171 recipe titles (f. 1r col. c - f. 1v col. c)
- 2) A collection of 169 culinary recipes (ff. 2r-21v)

#### B

- 3) An incomplete collection of 31 culinary recipes (ff. 22r-25v)

#### C

- 4) A collection of 291 medical recipes (ff. 30r-45v)
- 5) Medical miscellanea (ff. 45v-46r)
  - a) A medical maxim in Latin (f. 45v)
  - b) A list of oils, unguents and electuaries (ff. 45v-46r)
- 6) Miscellaneous medical recipes and herbal matter (f. 46r)
  - a) A medical apophthegm in Latin
  - b) A note on the virtues of vervain in Latin
  - c) A medical recipe in Middle English and some Latin

<sup>29</sup> A more detailed description of the contents of this MS is contained in appendix F and in the TEI header of the base data file for the MS included in appendix A.

- d) Herbal remedies in Latin
- e) A medical recipe in Latin
- f) A medical recipe in Middle English
- g) A medical recipe in Latin
- 7) Advice in Latin on diagnosis by uroscopy (f. 46v)
- 8) Two medical recipes in Middle English (f. 46v)
- D**
- 9) A trilingual herbal or *synonyma* (ff. 47r-54v)
- 10) Orthographical and grammatical advice (ff. 55r-56v)
  - a) Latin maxims of orthography and diction (ff. 55r-55v)
  - b) Latin grammatical mnemonics (ff. 55v-56v)
- E**
- 11) Seven bills of fare in Middle English (ff. 61)
- 12) A collection of 146 culinary and 20 veterinary recipes (ff. 61v-72v)
- F**
- 13) A collection of 137 medical recipes and two charms (ff. 73r-73v and 75r-81r)
- 14) Miscellaneous additions (f. 81r)
  - a) A medical recipe in Middle English
  - b) A medical recipe in Latin
  - c) Comments about Lent in Latin
  - d) Mainly a list of fishes in Latin with glosses in English (f. 81r)
- 15) Grammatical paradigms and examples in Latin (f. 81v)
- 16) An actual or exemplary letter in Latin (f. 82r)
- 17) Miscellaneous additions (f. 82v)
  - a) A pen trial in Latin, quoting Job 14:1
  - b) A pen trial in Latin
  - c) A pen trial in Latin
  - d) An incomplete copy of an exemplary Latin document (f. 82v)
- G**
- 18) A fragment of an epistle in Latin (ff. 74r)
- 19) A Latin treatise for a hermit or anchorite (ff. 74v and 83)
- 20) Miscellaneous additions (f. 83v)
  - a) A grammatical note with a mnemonic verse in Latin
  - b) A verbal exercise in Latin
  - c) Copy of the end of item 19 in a different hand
- H**
- 21) Medical recipes and a charm (ff. 84r-86r)
  - a) Eleven Middle English medical recipes (ff. 84r-86r)
  - b) A charm for the night evil in Middle English (f. 86r)
- 22) Four Middle English medical recipes (ff. 86v-87v)
- 23) Three Middle English medical recipes (f. 87v)
- 24) Medical recipes (ff. 89r-89v)
  - a) Eleven Middle English medical recipes with some Latin (ff. 89r-89v)
  - b) An incomplete medical recipe in Middle English (f. 89v)
- 25) 19 Middle English medical recipes (ff. 90r-91v)
- 26) Household and medical recipes in Middle English (f. 92r)
  - a) A method for catching coneyes (?)
  - b) A recipe for fish-bait
  - c) A medical recipe
  - d) A medical recipe

## I

- 27) A collection of over 18 Middle English medical recipes (ff. 94r-98r)
- 28) A medical recipe in Middle English (f. 98r)
- 29) Short verses, 11 in Latin, one in Middle English and one mixed (f. 100r)

**Hands used in the manuscript**

The different parts of this composite manuscript are written in a variety of different 15<sup>th</sup>- and early 16<sup>th</sup>-century hands, mostly anglicana-secretary hybrids of varying formality. Of the different scribal hands, the one responsible for the recipe collection (and the accompanying table of contents and bills of fare) is described in more detail, while the rest are given only a cursory description. The manuscript also contains a large number of additions and annotations in several hands, which significantly increases the total number of hands identifiable in the manuscript. The list of other hands identified in the manuscript is based on the observations made in Doyle (1992).

**Scribal hand** The original scribal hand for the recipe collection and the accompanying bills of fare and table of contents is a proficiently written anglicana-secretary hybrid written in brown ink with a minim height of around 2 mm. Doyle (1992: 3) dates the hand to the early 15<sup>th</sup> century and describes it as a "proficient" anglicana formata, which points toward the text having been produced by a professional scribe. There is, however, considerable variation in the formality of the hand within the text, which makes the qualification of the script as formata somewhat questionable for much of the text. This variation is not linear but fluctuates seemingly randomly, although there is a cluster of recipes exhibiting more formal characteristics between recipes 90 and 110. The ink is of a very light brown hue and thus clearly differentiated from the later additions, which for the most part are in a much blacker ink.

Although the script is in general closer to anglicana, the dominant form of *a* is a rather sharp-topped and angular single-compartment secretary *a*, but rounder variants also appear, as do some examples of the two-compartment anglicana *a*. The letter *d* is of the typical, rather rounded, anglicana variety with a looped ascender tilted to the left. The dominant form of *e* is the reverse Anglicana form, although the lobed secretary or gothic *e* also occurs frequently. The *f* is typically anglicana with its slender and upright descender, and *g* also appears in its usual anglicana form. In addition to the normal minim *i*, often marked as such by a curved hairline stroke above it—especially in words containing combinations of consecutive minims—the long form is used as a prefix. The long anglicana form of *r*—which has already lost its shoulder stroke—is the predominant type, with both the short secretary *r* and the 2-shaped ligatured *r* appearing intermittently. The *s* appears in all its variants (long, round and sigma-shaped), used rather indiscriminately with the long form being the dominant one. The letter *w* appears in its characteristic and relatively elaborate anglicana form.

**Other hands** In addition to the scribal hand used for the recipe collection, the following hands have been identified in the manuscript by Doyle (1992) and confirmed by the editor:

- 1) An anglicana hand in purplish black ink used for annotations and corrections in parts A-E of the manuscript (ff. 1-72). Although the formality of the hand varies, Doyle (1992: 13) describes it as expertly written and dates it to the middle of the 15<sup>th</sup> century. While very similar to the original hand used for part A of the manuscript, it is more cursive and written in a clearly different and much darker ink. It contains some secretary features, although it uses the two-compartment *a* and the anglicana-style short *s* more frequently than the original scribal hand. On the other hand, secretary *g*—which is not used by the original scribe—appears in some of the titles written in this hand.
- 2) A proficiently written early-15<sup>th</sup>-c. anglicana-secretary hybrid of varying formality in brown-black ink, which is the original scribal hand for item 3. This hand is very similar to the original scribal hand in shape, size and the extent of variation, and most belongs to the same scribe. The only differences between these two hands are the darker ink and the slightly larger and more formal recipe headings in item 3 which are absent from item 2.
- 3) A competent anglicana using the single-compartment secretary *a* written in brown ink and used for items 4-8 on ff. 30-46 (Doyle 1992).
- 4) An expertly written set secretary in pale greyish ink used for item 9 on ff. 47-54 (Doyle 1992).
- 5) A proficient leftwards-leaning secretary with anglicana *g*, written in dark brown ink and used for item 10 on ff. 55-56 (Doyle 1992).
- 6) A 15<sup>th</sup>-century hand of rather humanist duct used for two notes on hyssop on f. 52r and f. 53v (Doyle 1992).
- 7) A competently written mixture of anglicana and secretary, written in dark brown ink and used on lines 14-29 on f. 56v and for marginal additions on ff. 55-56 (Doyle 1992).
- 8) A competently written leftwards-leaning anglicana in brown ink, used for items 11 and 12 on ff. 61-72 (Doyle 1992).
- 9) The hand of a rubricator, in red pigment, that is used in parts A-E (ff. 1-72) both for decoration and for deletions and other emendations made in the text.
- 10) A competently written anglicana in brown ink, used for item 13 on ff. 73-76 and ff. 79-82 (Doyle 1992).
- 11) Another competently written anglicana with single-compartment *a* in brown ink, used for item 13 on ff. 77-79r (Doyle 1992).
- 12) A current hand used for item 14a on f. 81r (Doyle 1992).
- 13) A current hand used for item 14b on f. 81r (Doyle 1992).
- 14) A much less current hand used for item 14c on f. 81r (Doyle 1992).
- 15) A mid-15<sup>th</sup>-century hand used for item 15 on f. 81v (Doyle 1992).
- 16) The hand of a rubricator, in red pigment, used for highlighting and decoration in MS item 13 (ff. 73r-81r).
- 17) A small expertly written anglicana formata of documentary character used for the additions that make up item 17 on f. 82v (Doyle 1992).
- 18) A competently written leftwards-leaning anglicana in blackish ink, used on f. 74 and f. 83 (Doyle 1992).
- 19) The hand of a rubricator, in red pigment, used for highlighting and decorating the two leaves of a separate gathering misbound with part F (ff. 74 and 83).
- 20) An unevenly written anglicana in variable ink used for items 21 (ff. 84r-86r), 23 (f. 87v), 24b (f. 89v), 25 (ff. 90r-91v) and 28 (f. 98r) (Doyle 1992).
- 21) Another proficiently written anglicana, sometimes with single-compartment *a* and final secretary *s*, written in grey ink and used on ff. 86v-87v (Doyle 1992).
- 22) A competently written early anglicana (dated by Doyle (1992) to s. xiv<sup>2</sup>) with a rather squat aspect, written in pale brown ink and used on f. 89 (Doyle 1992).
- 23) A 15<sup>th</sup>- or 16<sup>th</sup>-century hand used for the household recipes on f. 92r (MS item 26a-c) (Doyle 1992).



- 24) The hand of a rubricator, in red pigment, used highlighting and decorating part H of the manuscript (ff. 84r-92r).
- 25) A proficiently written anglicana with single-compartment *a*, written in grey ink and used on ff. 94-100r (Doyle 1992).
- 26) The hand of a rubricator, in red pigment, used for highlighting and decoration in MS item 27 (ff. 94r-98r).
- 27) An early to mid-15<sup>th</sup>-century hand in ink used on
- 28) f. 100v to add some pen-trials, which has now faded very badly (Doyle 1992).
- 29) A rather cursive 16<sup>th</sup>-century hand used for items 20c (f. 83v), 26d (f. 92r) and 29 (f. 100), which mention the name John Bradfurth and the year 1542 (Doyle 1992).
- 30) A 15<sup>th</sup>- or 16<sup>th</sup>-century anglicana-secretary hybrid in black ink that has added the phrase "Souerand I ... lord haue mercy upon vs" on f. 12r.
- 31) A 16<sup>th</sup>-century cursive hand in black ink that added the phrase "in the name of god amen in this year of owere lord god a 155" in the outer margin of f. 4r.
- 32) A 16<sup>th</sup>-century hand that has added the phrase "god grant vs peaces the lord be wyth you" on f. 72v (Doyle 1992).
- 33) A mid- to late-16<sup>th</sup>-century hand in pale brown ink, used on f. 81.
- 34) A mid-16<sup>th</sup>-century cursive hand that added the phrase "Holy gos[t] haue mercy a ponvs myserable synneris John nycols boke" on f. 92r This hand is similar to hand 29 but written with a narrower nib.
- 35) An early 16<sup>th</sup>-century hand that added three recipes on f. 88r.
- 36) A mid-17<sup>th</sup>-century hand in black ink, belonging to George Davenport (c. 1631–1677), a chaplain of Bishop Cosin, which has added a table of contents for the manuscript and the name of its owner on ff. ii v-1r (Doyle 1992).
- 37) An early 18<sup>th</sup>-century hand in black ink, belonging to Thomas Rud, the librarian of Cosin Library c. 1720, and used for an ex libris inscription on f. 1r (Doyle 1992).
- 38) A late 18<sup>th</sup>-century hand in black ink, belonging to Robert Harrison, librarian of Cosin Library during the late 18<sup>th</sup> century, and used for pagination of parts A, B (ff. 1-25) and E (ff. 61-72) of the manuscript (Doyle 1992).
- 39) A late 18<sup>th</sup>-century hand in pencil, belonging to Robert Harrison, librarian of Cosin Library during the late 18<sup>th</sup> century, and used for the recipe numbers in parts A, B ff. 1-25 (recipe numbers) and E ff. 61-72 (recipe numbers) of the manuscript (Doyle 1992).
- 40) A modern hand in pencil, belonging to Dr. I. A. Doyle and used for foliating the entire manuscript c. 1951 (Doyle 1992).

## Decoration

The different parts of the manuscript, having been produced separately and only later collected together, vary in their extent of decoration. While it is impossible to show conclusively, the rubrication in parts A-E of the manuscript is so similar in style that it seems very possible that they were rubricated together, indicating an early association between them (see *Origins and provenance* below). In the following list, the decoration applied to the manuscript is described separately for each manuscript part:<sup>30</sup>

- A) In MS item 2, which is the recipe collection edited here, all of the recipes are preceded by a large (2-3 lines in height) paraph drawn in red pigment. A similar paraph has also been drawn next to the middle part of recipe number 101, the rubricator apparently being misled by the line-initial word "Take"

<sup>30</sup> The descriptions of parts C-I of the manuscript is based on Doyle (1992) and the editor's own cursory examination of those parts of the manuscript.

in the middle of the recipe. The beginnings of the bills of fare and table of contents preceding the recipe collection (MS items 1a, 1b and 1c) are indicated by large uncial capitals drawn in red pigment, the first of them being four lines in height and the other two being two lines in height. Almost all of the initial letters of the names of dishes in the bills of fare and the table of contents are highlighted by a vertical stroke of red pigment.

- B) The recipes and recipe titles in MS item 3 (ff. 22r-25v) are preceded by coloured paraphs drawn alternately in red and blue, the paraph of each recipe title being in the same colour as that of the previous recipe. This scheme, however, is realized only partially, the blue paraphs being drawn only on f. 22r, after which only every other title and recipe—following the scheme mentioned—have the paraph (in red). Capital letters and some punctuation in the recipes are also highlighted in red. Two large uncial initials on f. 25r signify the beginning of a new recipe in the middle of a text line along with two empty spaces, supposedly left for similar initials to be drawn in blue.
- C) Manuscript items 4-8 are decorated and organized using paraphs and double virgules drawn in red pigment. Titles of recipes have been underlined first in ink and then painted over with red pigment, along with braces drawn in red pigment on f. 46 and boxing of titles. Initial to item 4 is a 2-line letter drawn in red pigment.
- D) Capital letters in the trilingual herbal (MS item 9, ff. 47-54) are highlighted with red pigment and headings are framed in red pigment painted over ink. The first page, f. 47r, contains braces first drawn in ink and then painted over in red pigment, as well as red paraphs. In MS item 10 (ff. 55-56), capital letters are likewise highlighted with red pigment and f. 55r contains red paraphs and f. 56v has braces in red drawn over ink.
- E) In the recipe collection and the accompanying bills of fare (MS items 11-12, ff. 61r-72v), capital letters are highlighted with red pigment and recipe headings are framed in red. There are also a few red paraphs and the initial to MS item 11 on f. 61r is a 2-line letter in plain red.
- F) In MS item 13 (ff. 73r-81r, excluding f. 74) capitals are highlighted with red pigment, and red virgules and paraphs are used to structure the text. Underlining and boxing in red pigment are used for highlighting recipe titles and red line fillers are used to fill up incomplete lines at the ends of recipes.
- G) On the single bifolium from another manuscript that has been intruded into part F of the manuscript containing MS items 18, 19 and 20 (ff. 74 and 83), capital letters are highlighted with red pigment, red virgules and paraphs are used for structuring the text, and underlining in red ink is used for highlighting.
- H) In all of the items in part H of the manuscript (MS items 21-26, ff. 84r-92r), capital letters are highlighted with red pigment, and red virgules and paraphs are used to structure the text. Underlining in red pigment are used for highlighting recipe titles and red line fillers are used to fill up incomplete lines. F. 90v also contains some paraphs in blue pigment.
- I) In MS items 27-29 (ff. 94-100r), headings have been written in red pigment.

### Later additions

A large number of additions, annotations and corrections have been made to the different parts of the manuscript. The most significant of these are listed below, their order reflecting primarily the chronological order (by century) in which they were most likely added and secondarily their location in the manuscript:

- **15<sup>th</sup> century:**
  - Corrections, additions and annotations have been made to the text of parts A-E of the manuscript (ff. 1-72) by a mid-15<sup>th</sup>-century annotator in hand 1.
  - Headings have been added to recipes 11, 17, 20-41, 47-50, 52-66, 68-71, 74-76, 78-88, 90-93, 97-104, 107-8, 111-2, 114-7, 119-20, 123-7, 129-41, 143-8, 150-3 and 155-67 of item 2 (ff. 3-21) by the same mid-15<sup>th</sup>-century annotator (hand 1).
  - On each of the leaves of item 4, up to f. 33r (i.e. ff. 30-33r), a contemporary hand (possibly hand 1) has added sequences of letters from the beginning of the alphabet in the right margin, possibly for indexing. The rest of the pages are not indexed, however.
  - Corrections and notes by various unknown (probably 15<sup>th</sup>-century) hands have been made in part C of the manuscript (ff. 30-46).
  - Notes on hyssop in Latin have been added onto ff. 52r and 53v by a 15<sup>th</sup>-century hand of rather humanist duct (hand 6) (Doyle 1992).
  - Many recipes in item 12 (ff. 61v-72) have been marked *F* or *extra* in the margin, most likely in the hand of the 15<sup>th</sup>-century annotator (hand 1).
  - A number of pen-trials in an early or mid-15<sup>th</sup>-century hand (hand 27) have been made on f. 100v, an originally blank and unruled page. They are now very faded and only legible by ultra-violet light. According to Doyle (1992), some of them are ill-spelt, e.g. “Domus noster qui in est”, “Fuit omo missys adeo qui” [John 1:6]. Also includes two fragmentary quotations in English: “Than saw I noman at wold” and, based on Chaucer’s *Troilus and Criseyde* (Benson 1988: 1106-7), “So fer forþ am I broght in louys daunce | That endles I hop all way a bovt”.
- **16<sup>th</sup> century:**
  - The phrase “in the name of god amen in this yere of owere lord god a 155” has been added in the outer margin of f. 4r in a cursive 16<sup>th</sup>-century hand (hand 31).
  - The phrases “Sou<sup>e</sup> and I” and “lord haue mercy upon vs” have been added on f. 12r at the top of the page and between recipes 85 and 86 in a 15<sup>th</sup>- or 16<sup>th</sup>-century hand (hand 30).
  - The fragment of a leaf between ff. 45 and 46 contains the ends of 6 lines of text in an unknown mid-16<sup>th</sup>-century hand.
  - The phrase “god grant vs peaces the lord be wyth you” has been added on f. 72v in a 16<sup>th</sup>-century hand (hand 32) (Doyle 1992).
  - Two recipes have been added onto f. 81 in a mid- to late-16<sup>th</sup>-century hand in pale brown ink (hand 33).
  - A pen-trial in an unknown mid- to late-16<sup>th</sup>-century hand has been made on f. 84r.

- Three recipes have been added onto f. 88v in an early 16<sup>th</sup>-century hand (hand 35).
- The phrase “Holy gos[t] have mercy a ponvs myserable synneris John nycols boke” has been added onto f. 92r after item 26d in a mid-16<sup>th</sup>-century cursive hand (hand 34) (Doyle 1992).
- **17<sup>th</sup> century:**
- A list of contents by George Davenport (c. 1631–1677) has been added onto f. iiv, registering the following five heads, most likely referring to items 1-3, 4, 9, 11-12 and 13, respectively:
  - 1) Purveance for the feste for the King with the Lord Spenser, and receipts to make dishes.
  - 2) Gode medycines of Gode Leches.
  - 3) Nomina herbarum iueta Alphabetum.
  - 4) Receipts to make dishes.
  - 5) Medicamina, carmina, unguenta & c.
- An inscription of ownership, consisting of his name and the date 1652 has been added to the top margin of f. 1r by Davenport.
- **18<sup>th</sup> century:**
- Pagination has been added onto ff. 1–25 and 61–72 by Robert Harrison (1714/15–1802), a natural philosopher and linguist who served as the librarian for Bishop Cosin’s library at Durham Cathedral from about 1779 to 1796 (Doyle 1992).
- Continuous numbering (1-208) has also been added to the recipes of manuscript items 2 and 3 (ff. 2-25) in pencil by Harrison, with an unexplained (most likely erroneous) leap from 160 to 169.
- Recipe numbers have been added in ink (hand 38) to the margins of MS item 12 (ff. 61v-72) by Robert Harrison, apparently from the Pegge (1780) edition of the *Forme of Cury* or its reproduction by Warner (1791).
- **20<sup>th</sup> century:**
- Foliation has been added in pencil throughout the manuscript by Dr. Ian A. Doyle c. 1951.

### Origins and provenance

According to Doyle (1992: 13), the various parts of the manuscript have been written in England, although it is unlikely that they all originate in the same place. Items 13 and 16 may point to a connection with Merton College Oxford, and item 16 is a copy of a letter by a clerk with Kentish connections (Doyle 1992: 13). Otherwise the origin of the various parts of the manuscript is unknown. There is very little internal evidence for the dating of the manuscript, but if we accept the first bill of fare (MS item 1a) to refer to a feast “given by Thomas baron Despenser between 1394 when he had livery of his lands and 1397 when he was made earl of Gloucester”, it sets a *terminus post quem* of 1394 for the bills of fare and the associated recipe collection. The earliest hand identified in the manuscript by Doyle (1992) is hand 22, which he dates to the second quarter of the 14<sup>th</sup> century, but most of the scribal hands identified in the manuscript would seem to date from

around the middle of the 15<sup>th</sup> century. The watermarks of the papers used for the different parts of the manuscript would seem to have been in use from the late 14<sup>th</sup> to the early 15<sup>th</sup> century. In a brief mention of parts A and B of this manuscript, Hieatt and Butler (1985) remark that this section of the MS “appears to have been written early in the fifteenth century” (19).

Parts A-E were probably together already by the mid-15<sup>th</sup> century, when a rather current anglicana hand with some secretary features (hand 1) added titles to ff. 3–21, along with numerous corrections and notes to all of the aforementioned parts. Also the similarity of rubrication in the first five parts (similar red paraphs and vertical strokes on capitals in all of them, similar slanted boxes for titles in A, D and E) would seem to point towards their early association with each other. Based on a mention of his name and the date 1542 in MS item 29g, at least part of the manuscript probably belonged around this time to John Bradffurth, vicar of Lillington, who also made additions to several parts of the manuscript (see hand 29). The text “Holy gos[t] have mercy a ponvs myserable synneris John nycols boke” written on f. 92r in an italic hand looks like an owner’s inscription, and would place at least a part of the manuscript into the hands of someone called John Nichols, most likely sometime in the 16<sup>th</sup> century, although it is certain whether this would be before or after Bradffurth’s possession of the manuscript. Various other unidentified 16<sup>th</sup>-century hands have also made annotations on ff. 4r and 72r and onto the fragment of leaf between ff. 45 and 46, indicating that the manuscript was being actively read and used during this period.

In 1652 the manuscript was apparently acquired by George Davenport (c. 1631–1677)—a chaplain of John Cosin, Bishop of Durham (1595–1672)—who had it bound in its current binding and added a table of contents on the verso of f. ii, as well as his name with the date 1652 at the top of f. 1r (Doyle 1992: 1; Doyle 2004: 37). Sometime between its founding in 1669 and Davenport’s death in c. 1677 the manuscript became a part of the library of Bishop Cosin, also known as *Bibliotheca Episcopalis Dunelmensis* and founded by the Bishop of Durham, John Cosin (1595–1672), as an endowed public library for local clergy and people of scholarly interests (Cosin MSS). The ex libris inscription added above Davenport’s name is in the hand of Thomas Rud, the librarian of Cosin Library c. 1720 (Doyle 2006: pers. comm.), and must thus have been added at that time, perhaps when the manuscript was catalogued or otherwise processed. The manuscript ended up in the custody of the University of Durham as it became the trustee of the Cosin library in 1937, having had use of the Palace Green building housing the collection already from 1834.

### 9.3 Relationships to other collections

A manuscript that is unparalleled is, of course, always going to be harder to edit than one for which there are other witnesses that may offer helpful variants. And whether there are parallel Middle English manuscripts or not, parallels in French recipes can offer truly enlightening help. And not only *French* recipes: the more the editor knows about Continental (and, indeed, Arabic) recipe collections, the better.

(Hieatt 1998a: 137)

Although a thorough and systematic examination of the relationships between recipes in the different collections and families listed in subsection 8.1.1, not to mention the various foreign collections identified in Hieatt, Lambert et al. (1992) would require not only the transcriptions of previously unprinted recipe collections that Hieatt (1996: 54) calls for, but spelling-normalized and preferably semantically tagged electronic versions of these transcriptions, this section will nevertheless present some initial observations about the relationship of the *Potage Dyvers* family to other English and continental recipe collections.<sup>31</sup> As was pointed out above, the fact that the *Potage Dyvers* family is defined by the significant number of recipes shared between its members does not mean that its component recipes could not also be included in other collections that are nevertheless too different on the whole to be considered members of the same family. The conventional and arbitrary identity of the family also means that it is also possible that there exist further collections that have as good a claim to a membership in the *Potage Dyvers* family as the six versions edited here. Fortunately, the extensibility of digital editions—discussed in subsection 4.2.1—means that such newly discovered members can easily be edited as extensions to the present edition. Obviously, also recipes shared between collections of different families that are edited using principles similar to the ones described here can—and should—be linked together to form a network whose interconnections and overlaps can subsequently be studied in order to chart the transmission, selection and adaptation of recipes and the culinary inventions described by them over the centuries.

### 9.3.1 English collections<sup>32</sup>

Although recipes with the same or similar titles occur in most of the major Late Middle English collections, the recipes themselves often bear no resemblance with each other, frequently differing even in their basic ingredients. As an example, Hieatt and Butler (1985: 9–10) mention the originally Anglo-Norman dish known as ‘Mawmene’. While a large number of manuscripts from the early 14<sup>th</sup> to the late 15<sup>th</sup> century<sup>33</sup>—including all of the *PD* versions edited here (recipe *PD* 162)—contain a recipe by this name, most of the recipes are clearly different from each other, later ones being longer and more complicated than the earlier ones, to the extent that over the space of a century, even the dish itself often becomes unrecognisable. Of the ‘major’ families listed in subsection 8.1.1, none of the families originating in the 14<sup>th</sup>-century and edited by Hieatt and Butler (1985) (i.e. *Diversa*

<sup>31</sup> The fact that the manuscript transcriptions in this edition have been normalized and annotated with information about their basic word class will allow e.g. the comparison of process verb and ingredient noun sequences between recipes of different collections in order to discover general similarities in the processes described by recipes. Also the automated semantic tagging of the texts using new resources such as the *Historical Thesaurus of English* is being investigated as a postdoctoral project, building on the work done and the tools developed in the University of Glasgow (see e.g. Anderson et al. 2011). This, together with the planned creation of a larger corpus of historical English recipes would create a resource allowing the kind of comparative research envisioned by Hieatt (1996: 54) to be carried out using modern data-mining technology.

<sup>32</sup> Since the geographical origin of recipe collections or even individual manuscript versions of them is often extremely uncertain, the term *English* is here used in the loose sense of collections that have been written in Middle English or Anglo-Norman and most likely used in the British Isles.

<sup>33</sup> Hieatt and Butler mention examples from BL MS Additional 46919 (c. 1325), Bodleian MS Douce 257 (1381), BL MS Harley 5016 (c. 1425) and the MS H4016 version of *PD*.

*Cibaria*, *Diversa Servicia*, *Utilis Coquinario*, and *Forme of Cury*) seem to contain close parallels of the recipes with the collections edited here, although they do contain recipes for many of the same dishes. The 15<sup>th</sup>-century families, however, are a different story: not only do both of the other major 15<sup>th</sup>-century families listed by Hieatt (1992) share a number of recipes with the *PD* family, there also survive a number of individual manuscript collections—discussed below—which contain a significant number of recipes also included in a variable number of the *PD* versions.

### *An Ordinance of Pottage*

Of the other 15<sup>th</sup>-century families, the one edited by Hieatt (1988) as *An Ordinance of Pottage* (with Yale University Library MS Beinecke 163 as the base manuscript) shares around 30 recipes with many of the *PD* versions. These recipes, consist of a sequence of recipes for roasting different kinds of birds and a number of miscellaneous recipes scattered throughout the collection. The majority of these recipes—including the sequence of bird recipes—are shared between MSS Ad, D, C and H4016, although some of them belong to the ‘core’ group shared by all six of the versions (see chapter 13).<sup>34</sup> In addition to these recipes that fulfil the definition of a parallel recipe, there is also a very large number of recipes that are for the same basic dish type as a recipe in the *PD* but are clearly different recipes, describing a significantly different procedure or ingredients for the dish.

### *A Noble Boke off Cookry*

The last of the 15<sup>th</sup>-century collections listed in subsection 8.1.1, *A Noble Boke off Cookry*, edited by Napier (1882) from MS Holkham 674, which shows considerable similarities with *An Ordinance of Pottage*, also shares at least 34 recipes with the *PD* manuscripts. As with the recipes shared with *An Ordinance of Pottage*, the majority of the recipes occurring both in *PD* and the Holkham MS are shared between MSS Ad, D, C and H4016 but not MSS As, H279, although *NBC* also contains some recipes shared by all of the *PD* MSS, and at least one which occurs only in MSS As and H279.<sup>35</sup> The close resemblance between the *OP* and *NBC* families is apparent

<sup>34</sup> Recipes shared between *An Ordinance of Pottage* (*OP*) and all six of the *PD* versions include PD 27 (OP 105), PD 36 (OP 166), PD 48 (OP 65), PD 52 (OP 177), PD 57 (OP 87) and PD 72 (OP 131), while recipes shared between *OP* and MSS Ad, D, C and H4016 include PD 86 (OP 168), PD 105 (OP 132), PD 106 (OP 140), PD 107 (OP 141), PD 108 (OP 142), PD 109 (OP 143), PD 110 (OP 145), PD 111 (OP 146), PD 112 (OP 147), PD 113 (OP 148), PD 114 (OP 149), PD 115 (OP 144), PD 116 (OP 150), PD 118 (OP 152), PD 119 (OP 153), PD 120 (OP 154), PD 121 (OP 155), PD 123 (OP 157), PD 127 (OP 162), PD 133 (OP 169), PD 134 (OP 170), PD 138 (OP 171), PD 148 (OP 175), PD 167 (OP 63) and PD 168 (OP 64).

<sup>35</sup> The recipes shared between *NBC* and all six of the *PD* versions include PD 27 (NBC 43), PD 49 (NBC 178), PD 64 (NBC 139), PD 68 (NBC 133), PD 72 (NBC 72) and PD 83 (NBC 129), while recipes shared between *NBC* and MSS Ad, D, C and H4016 include PD 75 (NBC 134), PD 78 (NBC 131), PD 79 (NBC 130), PD 86 (NBC 102), PD 106 (NBC 77), PD 108 (NBC 78), PD 109 (NBC 79), PD 115 (NBC 80), PD 107 (NBC 81), PD 110 (NBC 82), PD 111 (NBC 83), PD 112 (NBC 84), PD 113 (NBC 85), PD 114 (NBC 86), PD 118 (NBC 89), PD 119 (NBC 90), PD 120 (NBC 91), PD 121 (NBC 92), PD 122 (NBC 93), PD 127 (NBC 97), PD 133 (NBC 104), PD 134 (NBC 105), PD 138 (NBC 106), PD 167 (NBC 174) and PD 175 (NBC 151). In addition to these, *NBC* also includes four recipes for sauces which are shared by all of the *PD* MSS except for MS H279—PD 98 (NBC 125), PD 96 (NBC 126), PD 95 (NBC 127), PD 94 (NBC 128)—and at least one recipe—PD 228 (NBC 231)—that occurs only in MS H279. The recipes

in the fact that many of the recipes shared between them and the *PD* family—most notably the sequence of bird recipes mentioned above—are the same ones that are also shared with the *OP*, as apparent from *Table 9.2*. In addition to these shared recipes, *NBC*—like *OP*—also contains a number of recipes that are not the same as those in the *PD* versions but clearly describe the same dishes.

#### London, British Library MS Harley 5401 (H5401)

Of the currently known recipe collections surviving in a single manuscript copy and containing recipes found in at least one of the *Potage Dyvers* versions, the first was described and edited by Hieatt (1996). This relatively short collection of 96 recipes is found in BL MS Harley 5401, a manuscript miscellany containing mostly medical material, and “has an unusual relationship to a remarkably large group of ‘sources’”, including the *Forme of Cury*, the *Diuersa Servisa*, and the *Potage Dyvers* families. While it contains a large number of the same recipe titles as the *Potage Dyvers*, the contents of the recipes themselves are often very different. Of the 96 recipes, seven seem to be parallels of recipes occurring in the *PD* family, most of them located near the end of the collection. Based on the remarks of Hieatt (1996: 66–9) and a brief examination of the recipes in the two collections, the *PD* recipes occurring in MS Harley 5401 include: PD 60 (H5401 84), PD 64 (H5401 85), PD 78 (H5401 87), PD 80 (H5401 88), PD 242 (H5401 53), PD 273 (H5401 16) and PD 310 (H5401 89). Within the *PD* family these recipes do not form a unified group and their distribution is an unusually irregular one; three of them (PD 60, 64 and 80) occur in all of the versions, one (PD 78) occurs in the closely related group made up of MSS Ad, D, C and H4016 (see chapter 13), while another three (PD 243, 274 and 311) occur only in MSS As, H279, indicating that they most likely originate in different sources.

#### Oxford, Bodleian Library MS e. Mus. 52 (M52)

Hieatt (2008: 45–61) describes the recipe collection contained in Bodleian MS e. Mus. 52—a 16<sup>th</sup>-century miscellany manuscript—as having been written in the 15<sup>th</sup> century based on its hand. The collection combines recipes occurring in a number of the families identified by Hieatt (1992) and listed in subsection 8.1.1, including the *Potage Dyvers*, the *Forme of Cury* (*FC*), and *An Ordinance of Pottage*, along with some recipes Hieatt reports not having encountered elsewhere. Of the 101 recipes—excluding the 26 confectionary recipes following the culinary ones—Hieatt indicates more than half (53) to be parallels of ones found in the *PD* versions edited by Austin (1888), while 11 occur in the *Forme of Cury* (*FC*), 4 in *An Ordinance of Pottage* and 33 do not occur elsewhere. The *PD* recipes Hieatt reports as occurring in this MS, ordered by their recipe references (see subsection 11.9.2), are:<sup>36</sup>

---

have not been numbered in Napier (1882), but the numbers used here are based on the order of the recipes in that edition.

<sup>36</sup> Unfortunately Hieatt (2008) does not print the MS e. Mus. 52 recipes that have parallels elsewhere in their entirety, which makes it impossible to evaluate the accuracy of these results.



Ref.	Normalised title	PD MSS	No. in <i>OP</i>	No. in <i>NBC</i>
PD 27	Doves Stewed	H279,Ad,As,D,C	43	105
PD 36	Stuffed Chicken	H4016,H279,Ad,As,D,C		166
PD 48	Brawn in Comfit	H4016,H279,Ad,As,D,C		65
PD 49	White Brawn	H4016,H279,Ad,As,D,C	178	
PD 52	Sops Chamberlain	H4016,H279,Ad,As,D,C		177
PD 57	Pears in Compost	H4016,H279,Ad,As,D,C		87
PD 64	Alayed sops	H4016,H279,Ad,As,D,C	139	
PD 68	Rabbit in Gele	H4016,H279,Ad,As,D,C	133	
PD 72	Pies of Paris	H4016,H279,Ad,As,D,C	72	131
PD 75	Pike in Galentine	H4016,Ad,D,C	134	
PD 78	Mussels in Broth	H4016,Ad,D,C	131	
PD 79	Mussels in the Shell	H4016,Ad,D,C	130	
PD 83	Garbage	H4016,H279,Ad,As,D,C	129	
PD 86	Breast of Mutton in Sauce	H4016,Ad,D,C	102	168
PD 94	Garlic Pevorade	H4016,Ad,As,D,C	128	
PD 95	Green Sauce	H4016,Ad,As,D,C	127	
PD 96	Ginger Sauce	H4016,Ad,As,D,C	126	
PD 98	Galentine	H4016,Ad,As,D,C	125	
PD 105	Brinews	H4016,H279,Ad,As,D,C		132
PD 106	Roast Swan	H4016,Ad,D,C	77	140
PD 107	Roast Crane	H4016,Ad,D,C	81	141
PD 108	Roast Pheasant	H4016,Ad,D,C	78	142
PD 109	Roast Partridge	H4016,Ad,D,C	79	143
PD 110	Roast Heron	H4016,Ad,D,C	82	145
PD 111	Roast Bittern	H4016,Ad,D,C	83	146
PD 112	Roast Egret	H4016,Ad,D,C	84	147
PD 113	Roast Curlew	H4016,Ad,D,C	85	148
PD 114	Roast Brewwe	H4016,D,C	86	149
PD 115	Roast Quail	H4016,Ad,D,C	80	144
PD 116	Rabbit	H4016,Ad,D,C		150
PD 118	Roast Duck	H4016,Ad,D,C	89	152
PD 119	Plover	H4016,Ad,D,C	90	153
PD 120	Snipe	H4016,Ad,D,C	91	154
PD 121	Roast Woodcock	H4016,Ad,D,C	92	155
PD 123	Roast Veal	H4016,D,C		157
PD 127	Fillets of Pork Endored	H4016,Ad,D,C	97	162
PD 133	Boiled Trout	H4016,Ad,D,C	104	169
PD 134	Boiled Lobster	H4016,Ad,D,C	105	170
PD 138	Roast Bream in Sauce	H4016,Ad,D,C	106	171
PD 148	Boiled Sturgeon	H4016,Ad,D,C		175
PD 167	Roast Turbot in Sauce	H4016,Ad,D,C	174	63
PD 168	Roast Salmon in Sauce	H4016,Ad,D,C		64
PD 227	Chicken in Gauncele	H279,As	231	

**Table 9.2:** *The recipes shared by the PD manuscripts with the An Ordinance of Pottage and A Noble Boke off Cookry families.*

PD 1 (M52 58), PD 2 (M52 59), PD 3 (M52 60), PD 4 (M52 61), PD 5 (M52 62), PD 6 (M52 12, 63), PD 7 (M52 41), PD 8 (M52 13), PD 9 (M52 64), PD 10 (M52 65), PD 11 (M52 49), PD 12 (M52 50), PD 14 (M52 52), PD 15 (M52 53), PD 16 (M52 54), PD 17 (M52 55), PD 18 (M52 56), PD 28 (M52 1), PD 29 (M52 2), PD 30 (M52 3), PD 31 (M52 89), PD 32 (M52 90), PD 33 (M52 91), PD 46 (M52 14), PD 47 (M52 4), PD 48 (M52 5), PD 49 (M52 6), PD 52 (M52 15), PD 54 (M52 10), PD 55 (M52 11), PD 56 (M52 7), PD 57 (M52 8), PD 59 (M52 9), PD 60 (M52 80), PD 61 (M52 74), PD 62 (M52 73), PD 63 (M52 72), PD 64 (M52 71), PD 65 (M52 70), PD 66 (M52 69), PD 67 (M52 68), PD 69 (M52 48), PD 102 (M52 19), PD 105 (M52 18), PD 164 (M52 86), PD 166 (M52 20), PD 172 (M52 16), PD 173 (M52 17), PD 276 (M52 21), PD 366 (M52 83, 94) and PD 367 (M52 95).

The vast majority (41) of these recipes are ones which belong to the ‘core’ group of *PD* recipes that occur in five or six of the MS versions. As can be seen from this list, there are several groups of recipes that occur as consecutive series in both MSS. The longest of these is made up of recipes PD 1–18, which occur as a consecutive series at the beginning of MSS Ad and D and scattered over the other versions.<sup>37</sup> In MS e Mus. 52, these recipes occur as a consecutive series near the middle of the collection, with PD 11–18 occurring first as a series, followed by PD 1–10. Another interesting group is made up of recipes PD 60–67, which occur in inverted order in MS M52.<sup>38</sup> Based on the ordering of these sequences (and several shorter ones of two or three recipes), the closest relations of this MS among the *PD* versions are MSS Ad and D (see chapter 13). While the fact that this collection shares more than half of its recipes with the *PD* family would seem to speak for its inclusion in the family, the fact that it only contains less than a third of the recipes in even the smallest *PD* version and less than half of the ‘core’ recipes, it is best seen as a subselection or an excerpt of the *PD* family, combined with smaller subselections of recipes from other sources.

#### Oxford, Bodleian MS Rawlinson D 1222 (R)

This manuscript is a relatively long MS of 76 folia, in which Hieatt (2008: 62) counts a total of 281 recipes.<sup>39</sup> According to Hieatt (2008: 62–4), the largest number of recipes with an identifiable source in this manuscript (total of 104) come from *An Ordinance of Pottage*, while 68 of the recipes occur in at least one of the *PD* versions edited here, and 6 occur in *A Noble Boke off Cookry*.<sup>40</sup> Some of the recipes (6) are “obviously versions of the fourteenth-century recipes” printed in Hieatt and Butler (1985) (from the *DS* and *FC* families as well as the miscellaneous recipes labeled “Goud Kokery” by Hieatt and Butler), while 95 of the recipes are not attested in any other collection. The recipes shared by the *PD* family with this manuscript include, in the order of their recipe references in the present edition:

<sup>37</sup> Apart from PD 1 and 13 which do not occur in MS H4016, PD 8 which does not occur in MS C, and PD 9 which does not occur in MSS As and H279, these recipes occur in all of the *PD* versions.

<sup>38</sup> This inverted order is not seen in any of the *PD* versions.

<sup>39</sup> The collection was originally even longer, as it seems to be missing at least five folia at the beginning, with the first surviving recipe missing its beginning.

<sup>40</sup> Four of these six are also found in *OP*.

PD 3 (R 128), PD 4 (R 125), PD 6 (R 272), PD 11 (R 129), PD 16 (R 205), PD 17 (R 217), PD 18 (R 130), PD 28 (R 122), PD 29 (R 132), PD 32 (R 133), PD 33 (R 134), PD 34 (R 135), PD 35 (R 136), PD 37 (R 219), PD 38 (R 137), PD 39 (R 138), PD 40 (R 210), PD 42 (R 139), PD 44 (R 140), PD 47 (R 123), PD 51 (R 141), PD 52 (R 142), PD 53 (R 143), PD 56 (R 124), PD 61 (R 274), PD 62 (R 275), PD 64 (R 144), PD 69 (R 145), PD 74 (R 276), PD 77 (R 146), PD 78 (R 147), PD 80 (R 148), PD 83 (R 149), PD 85 (R 220), PD 89 (R 150), PD 91 (R 211), PD 104 (R 151), PD 105 (R 152), PD 130 (R 153), PD 131 (R 56), PD 132 (R 57), PD 135 (R 60), PD 136 (R 61), PD 137 (R 62), PD 139 (R 65), PD 140 (R 66), PD 142 (R 69), PD 143 (R 70), PD 144 (R 71), PD 145 (R 72), PD 146 (R 73), PD 147 (R 68), PD 152 (R 277), PD 153 (R 278), PD 158 (R 154), PD 159 (R 155), PD 161 (R 156), PD 165 (R 157), PD 169 (R 158), PD 171 (R 159), PD 174 (R 279), PD 177 (R 160), PD 179 (R 209), PD 180 (R 280), PD 181 (R 161), PD 184 (R 63), PD 252 (R 173) and PD 356 (R 188).

While organizational principle of the Rawlinson MS is completely different from that of any *PD* version, the recipes being divided into specific categories based on their method of preparation, such as “fish broiled or roasted”, “rosted metes” or “gely” (Hieatt 2008: 62), this MS does—like MS e. Mus. 52 above—share some ordered sequences of recipes with MSS Ad and D, which constitute the closest parallels of this MS in the *PD* family (Hieatt 2008: 64). The most prominent of these sequences is the sequence of fish recipes (PD 131–2, 135–7, 184, 139–40, 147, 142–6) which also occurs as a unit in MSS Ad, D, C and H4016 (see chapter 13). More than half (38) of the recipes shared between the *PD* family and MS Rawlinson D 1222 belong to the ‘core’ group shared by all or all but one versions. Of the rest, the clear majority (26, including the above mentioned group of fish recipes) are ones shared between the four *PD* manuscripts mentioned. In addition to these, there is a single recipe (PD 252) that occurs only in MSS Ad and H279, one (PD 356) that is unique to MS H4016 and one (PD 177) that is shared between MSS Ad, H279, Ad and D—an unusual combination as will become apparent in chapter 13. While MS Rawlinson D 1222 shares a large number of recipes with the *PD* family in absolute terms, they represent only a quarter of the total recipes in this collection, and cover less than half of the ‘core’ *PD* recipes, clearly distinguishing it from the *PD* family.

### London, British Library MS Sloane 1108 (SI)

Of the collections not traditionally included in the *Potage Dyvers* family, MS Sloane 1108 has the strongest claim to its membership. According to Hieatt (2008: 96), this early 15<sup>th</sup>-century collection contains 136 recipes, with at least one folio missing from between the first and second folia.<sup>41</sup> According to Hieatt (2008: 96), the recipes of this collection that are shared with the *Potage Dyvers* family are “probably often most close to the Douce 55 readings, but not in the Douce order” (Hieatt 2008: 96). The vast majority—98 of 136—of the recipes in this collection are ones shared with at least one of the *PD* versions.<sup>42</sup> These recipes, ordered by their reference number, are:

<sup>41</sup> Since the original numbering of the recipes in Roman numbers is thus inconsistent, the editorial numbering provided by Hieatt, reflecting the current state of the manuscript, is used instead.

<sup>42</sup> The the rest are either shared with one of the 14<sup>th</sup>-century collections edited in Hieatt and Butler (1985) (25) or unique recipes not attested anywhere else (11), except for two recipes that occur among the 25 recipes contained in Bodleian MS Laud 553 and edited in Austin (1888).

PD 1 (SI 17), PD 2 (SI 18), PD 3 (SI 19), PD 4 (SI 20), PD 5 (SI 22), PD 6 (SI 57), PD 7 (SI 80), PD 8 (SI 58, 85), PD 9 (SI 81), PD 10 (SI 82), PD 11 (SI 23), PD 12 (SI 24), PD 13 (SI 25), PD 14 (SI 59), PD 15 (SI 1), PD 16 (SI 55), PD 17 (SI 71), PD 18 (SI 2), PD 19 (SI 21), PD 20 (SI 26), PD 22 (SI 72), PD 23 (SI 56), PD 24 (SI 83), PD 25 (SI 5), PD 26 (SI 96), PD 30 (SI 78), PD 36 (SI 60), PD 38 (SI 3), PD 39 (SI 4), PD 41 (SI 73), PD 45 (SI 74), PD 47 (SI 6), PD 48 (SI 79), PD 49 (SI 61), PD 50 (SI 62), PD 51 (SI 7), PD 53 (SI 8), PD 54 (SI 9), PD 55 (SI 10), PD 56 (SI 11), PD 57 (SI 13), PD 60 (SI 92), PD 61 (SI 84), PD 63 (SI 86), PD 64 (SI 14), PD 65 (SI 15), PD 66 (SI 27), PD 67 (SI 28), PD 68 (SI 29), PD 69 (SI 30), PD 70 (SI 77), PD 73 (SI 76), PD 76 (SI 31), PD 77 (SI 32), PD 82 (SI 33), PD 83 (SI 34), PD 84 (SI 35), PD 87 (SI 63), PD 90 (SI 36), PD 92 (SI 65), PD 93 (SI 66), PD 94 (SI 67), PD 96 (SI 68), PD 98 (SI 69), PD 99 (SI 70), PD 102 (SI 93), PD 104 (SI 37), PD 105 (SI 38), PD 124 (SI 87), PD 125 (SI 95), PD 126 (SI 94), PD 151 (SI 39), PD 153 (SI 88), PD 154 (SI 89), PD 155 (SI 90), PD 156 (SI 91), PD 158 (SI 12), PD 159 (SI 16), PD 160 (SI 41), PD 161 (SI 42), PD 162 (SI 43), PD 164 (SI 44), PD 165 (SI 45), PD 166 (SI 40), PD 171 (SI 46), PD 172 (SI 48), PD 173 (SI 49), PD 174 (SI 52), PD 175 (SI 50), PD 176 (SI 47), PD 177 (SI 51), PD 178 (SI 53), PD 181 (SI 54), PD 183 (SI 64), PD 215 (SI 98), PD 232 (SI 116) and PD 242 (SI 121).

The majority of the *PD* recipes—62 of the 98—belong to the category of ‘core’ recipes. In addition to these, 30 recipes are shared by at least three of the closely related MSS Ad, D, C and H4016 (see section 13.1), of which 5 sauce recipes also occur in As. In addition to these relatively central ones, this manuscript also contains three recipes which are shared by MSS As, H279, Ad and D, and three more which are shared only by MSS As and H279.

In terms of its organization, the Sloane MS does not resemble any of the *PD* versions, although it does contain a large number of short sequences of two to four recipes occurring in the same order as in MSS Ad and D. While this MS does not contain either of the frequently occurring sequences for birds or fish, it does contain a sequence of six sauce recipes that occur in the same order in MSS Ad and D (and in almost the same order in all other versions except for MS H279). While this collection is still much smaller than any of the known *PD* versions, it nevertheless contains more than half of the ‘core’ recipes and shares just over half of all the recipes contained in MSS Ad and D, which means that it could well be considered a member of the *PD* family—or at least as an indication that the division of the 15<sup>th</sup>-century recipe collections into closed ‘families’ is an artificial and largely arbitrary categorization, and that the interrelationships obtaining between the various surviving collections are more complicated than a simple division into families would lead to understand.

### Aberystwyth, National Library of Wales MS Peniarth 394 D (P)

In contrast to the previous one, this manuscript—described by Hieatt (2008: 118–9)—contains only a minority—32 out of 285—of *PD* recipes, the majority (178) of the recipes being ones associated with the *Forme of Cury* family of collections and the rest being either shared with a number of different smaller collections or unique to this manuscript. According to Hieatt (2008: 118), who unfortunately does not include the content of the shared recipes and compares them only to Austin’s printed edition of MS Harley 4016, the recipes in the Peniarth MS are not always exact duplicates of that MS, but contain “small, relatively insignificant, differences” (Hieatt 2008: 118), and may in fact be closer to the versions contained in MSS Ad and D, both in terms of their content and ordering. As Hieatt notes, the recipes shared with the *PD* versions are scattered in groups throughout the collection and include the following recipes, ordered according to their recipe reference:

PD 73 (P 111), PD 75 (P 112), PD 78 (P 113), PD 80 (P 114), PD 82 (P 115), PD 91 (P 116), PD 130 (P 117), PD 133 (P 120), PD 134 (P 121), PD 135 (P 122), PD 136 (P 123), PD 137 (P 124), PD 139 (P 128), PD 142 (P 132), PD 143 (P 133), PD 145 (P 134), PD 146 (P 135), PD 147 (P 131), PD 151 (P 167), PD 152 (P 168), PD 153 (P 169), PD 154 (P 170), PD 155 (P 171), PD 156 (P 172), PD 157 (P 173), PD 163 (P 175), PD 165 (P 176), PD 167 (P 177), PD 168 (P 178), PD 169 (P 179), PD 170 (P 180), PD 181 (P 187) and PD 184 (P 127)

The majority of these recipes shared with the *PD* family comprise two separate series of fish recipes (P 111–7, 120–4, 127–8, 131–5 and 175–80, interrupted and extended by further fish recipes from other sources) that replicate not only most of the fish recipe sequence shared by MSS Ad, D, C and H4016 (see section 13.2), but also includes other fish and seafood recipes occurring elsewhere in the *PD* versions. While the majority of the recipes shared between the *PD* family and the Peniarth MS are ones occurring only in MSS Ad, D, C and H4016, this MS also contains 9 recipes that belong to the ‘core’ group, most of which (4) occur in the short section of miscellaneous fish day recipes (P 167–73) occurring between the two groups of fish recipes. Based on the selection of recipes, it would seem that in this MS, the recipes borrowed from the *FC* family have been supplemented by the extensive—and popular—selection of fish recipes shared not only by the four *PD* versions mentioned above but also by MS Rawlinson D 1222 described above.

#### Cambridge, Trinity College MS 0.1.13

This manuscript, described by Mooney (1995) and Hieatt (2008: 139), contains 102 recipes and constitutes a version of *An Ordinance of Pottage*, supplemented by a selection of recipes from other sources, including one shared by the *Potage Dyvers* family. This shared source has provided the first four recipes of the collection, which are parallels of the four first recipes in MSS Ad and D (PD 1–4), which also occur in all of the *Potage Dyvers* versions except for MS H4016 which contains only PD 2 and 4.

#### Oxford, Bodleian MS Rawlinson D 913

This manuscript consists of 34 fragments, of which one contains three whole recipes and three fragments, of which two whole recipes and two fragments are parallel versions of recipes shared either by all of the *PD* versions (PD 21, PD 22 and PD 172; recipes 4, 5 and 3 in this MS) or by MSS Ad, D, C and H4016 (PD 176; recipe 2 in this MS).<sup>43</sup>

### 9.3.2 Continental collections

Considering the cosmopolitan nature of upper-class late medieval society and the interrelationships between English and Continental culinary writing, apparent in the prevalence of French terminology and the use of closely related recipe titles on both sides of the channel, the lack of clear parallels between surviving Continental and English recipes is somewhat surprising and also somewhat unfortunate. As for example Keiser (1998c) has pointed out, a foreign language original for a group of Middle English recipes would provide “a very valuable tool to use in determining manuscript relations” (112). Unfortunately, no obvious continental

<sup>43</sup> PD 21–2 are not included in MS C in its current state, but based on the table of contents they were originally included on the folia that have subsequently been lost.

models for the recipes contained in the *Potage Dyvers* family of collections have been found. While this does not mean that they would not have been ultimately based on foreign models, it would seem to indicate that England had by the 15<sup>th</sup> century already developed its own distinctive version of the European culinary tradition. For example Scully (1988: 28), based on his extensive study of the textual tradition of the popular French recipe collection *Le Viandier de Taillevent*, notes that while the English cookery of the 15<sup>th</sup> century “may owe a general debt to France for much of its inspiration, no particular links with the *Viandier* or its long continental French tradition are evident” in the English collections.<sup>44</sup>

The extensive variation among the surviving versions of the *Potage Dyvers* (see section 13.4) implies an extensive transmission history in a Middle English context, meaning that any continental origins—should they exist—would have been buried under multiple layers of copying and adaptation in English. Thus the closest correspondence we can expect between the surviving versions of the *Potage Dyvers* family and any continental recipes exists only on the level of the dishes described, and even those are likely to differ significantly in their details. Furthermore—as was observed in chapter 8—the fluid, largely oral and cosmopolitan nature of European culinary culture means that even if such correspondences were to be found, they do not necessarily indicate any kind of textual relationship between the two recipes.

While it would be extremely interesting and useful from a culinary historical point of view to trace the movement of dishes and recipe variants on a pan-European level, the fact that they are written in different languages and are likely to have undergone significant changes in translation and transmission means that the comparison of these recipes will need to take place purely on the semantic level. In practice this means either an extremely laborious process of manual comparison—which becomes exponentially more complex with an increase in the number of recipes and collections to be compared—or the automatic comparison of semantic annotation describing the principal components of the recipes, i.e. the ingredients and operations described by them. While the former would in principle be possible for those recipe collections that have been edited in some form, the amount of manual labour is not only beyond the scope of this thesis but also not very cost-effective in any context. The latter option, which would allow for the algorithmic comparison of recipes for similarities in their culinary content, would require the semantic annotation of the recipes to be compared. While this kind of an annotated recipe database is not available at the moment, the editorial principles and practices described in this thesis—including the normalization of lexis and word-class disambiguation—do provide a solid foundation for the semi-automatic generation of such annotation, and for the future creation of such a database of not only Middle English but also of medieval French and Italian recipe collections. This kind of a database might one day provide the means for tracing the development and spread of European culinary culture.

<sup>44</sup> For the 14<sup>th</sup>-century recipe collections edited by Hieatt and Butler (1985), they still find some relatively close parallels in *Le Viandier de Taillevent* (subsequently edited by Scully 1988).

## Chapter 10

# Editorial principles

A strategy devised in the light of the materials, the media of presentation, and the primary audience sets priorities and guides design. Editions are problem-solving mechanisms; without problems to solve, new editions would not be needed. Explicit strategy is essential, because the problems of producing hypermedia editions have not been fully faced or solved. (Eaves 2006: 212)

Building on the theoretical background laid out in chapters 3 and 4, this chapter will introduce the editorial principles and practices used to produce the corpus-linguistically oriented digital documentary edition of the *Potage Dyvers* family of recipe collections that forms the core of this thesis. However, in addition to describing the present edition, both this description of the editorial principles and the technical documentation of annotation practices in chapter 11 are also intended to be programmatic in the sense of laying the basis for a set of practical guidelines for creating further digital editions for the purposes of corpus linguistic study. As such they constitute a modest attempt to answer the call made by Hockey already in 2004 for editorial projects that “concentrate more on the design and testing of [...] intellectual frameworks” for producing digital editions and on “the long-term implications of how these editions are being prepared”, instead of just on the “preparation of specific editions”, so that we “will be able to work toward a commonly accepted model that will last into the future” (Hockey 2004: 373).

In terms of the three-layered ontology of digital editions presented in chapter 4, *the edition* in this context refers exclusively to the *data archive* constituting a model of the edited document, apart from any specific processing engine or the editorial output produced by it.<sup>1</sup> The definition and documentation of the edition as a

---

<sup>1</sup> While this data archive is here seen to constitute the essence of the digital edition and to define its identity, its potential in terms of different editorial outputs is demonstrated by two different ‘print-friendly’ editorial outputs contained in appendices B and C, and a hypertext parallel reading edition of the six manuscripts contained in appendix D. All of these editorial outputs have been produced algorithmically by a series of XSLT scripts included and described in appendix E

data source rather than an editorial presentation is based on the observation, made in chapter 4, that the processing engine—and thus the editorial presentation—of a digital edition is not only likely to become obsolete long before its content but is also difficult to curate or archive. As a digital archive with extensively documented editorial principles, encoded using recognized standards like XML and the *TEI Guidelines*, the present edition has been designed to answer the call of Deegan (2006: 360-3) for digital editions that can easily be collected and preserved by libraries, and not only survive beyond the next five years but be able to make use of tools and interfaces developed long after their initial creation.<sup>2</sup>

In terms of the theoretical background presented above, the present edition is conceived as a *linguistically oriented digital documentary edition* of a *dispersed discourse colony* made up by the recipes contained in the six *Potage Dyvers* manuscript versions. The design principles of the edition, which have also guided the adaptation of the *TEI Guidelines* outlined in chapter 11 below, are based on the practical requirements of corpus linguistic research discussed above in subsection 4.3.3. This edition meets those requirements by

- 1) digitally encoding the whole of the textual content of each manuscript version separately as a faithful diplomatic transcription;
- 2) representing the visual and physical features of the textual artefact (layout, decoration and emphasis, damage, etc.) in the form of descriptive XML annotation integrated to the transcription;
- 3)
  - a) explicitly recording all instances of emendation in the original document and linking each span of text to the hand it was written in, and
  - b) annotating all editorial interpretations (such as the expansion of abbreviations) in a way that allows them to be removed without altering the original textual content;
- 4) storing the edition in the form of highly structured XML documents which can not only be computationally analysed using widely available and standardized technologies, but also transformed to various presentation or database formats to meet varying research requirements;
- 5) providing the text with an explicit *textual coordinate system* to which additional overlays of annotation can be linked; and
- 6) documenting in great detail the bibliographic details of the original documents, the editorial principles followed in the edition, and the annotations used in the edition.

In fulfilling the requirements set by the demands of corpus linguistic research, the present edition also follows the design principles set by the CES (EAGLES 2000) to guide the encoding and annotation of linguistic corpora, namely:

- 1) **Adequate coverage:** the standard should include unambiguous and uniform annotations for all those—and only those—features of the text that are relevant for language engineering applications;
- 2) **Consistency:** the features should be encoded in a consistent and logical way;

<sup>2</sup> An example of a similar approach is the *MENOTA* project, which maintains a strict separation between the archive versions of their transcriptions—maintained as XML documents—and the presentation versions—formatted as HTML or PDF documents—that are generated from the archive versions by XSLT stylesheets (Haugen 2004: 87).



- 3) **Recoverability**: it should be possible to separate what was in the original document from any annotation added by the corpus compilers using simple algorithms;
- 4) **Validatability**: the elements used should be specific enough and element hierarchies strict enough that structural constraints can be automatically validated;
- 5) **Capturability**: the minimum level of initial annotation performed during text capture should not be too costly, and it should be possible to refine it at a later stage;
- 6) **Processability**: it should be possible to automatically convert encoded texts into other encoding formats and to efficiently select texts based on user-specified criteria;
- 7) **Extensibility**: it should be possible to add new elements to the standard in a controlled and predictable way in case there arises a need to encode new features;
- 8) **Compactness**: annotation should be as compact as possible without sacrificing processability; and
- 9) **Readability**: the annotated text should remain human-readable.

These solutions mean that the edition follows to the principles of *flexibility*, *transparency* and *expandability*, outlined in Honkapohja, Kaislaniemi and Marttila (2009), and not only fulfils the requirements of modern historical linguistics but also encourages

- 1) the development of tools and user interfaces through strictly defined and well-documented annotation practices,
- 2) formulation of novel research questions by representing both the text and its physical manuscript context in an easily manipulable and queryable form,
- 3) sharing and cumulative production of knowledge by providing a mechanism for linking the results of linguistic or textual analyses to the base transcription and for sharing them with other users of the edition.

The first section of this chapter will explain the conceptual structure of the edition, seen as a *data archive*, in terms of the different types of data and metadata included in the edition and their relationships to each other and to the original document, as well as the methods used concurrently annotate multiple ontological structures involving overlapping hierarchies. The second section explains the treatment of the textual content of the document in the edition, including the delimitation of the textual content included in the edition, the transcription of letterforms, the handling of capitalisation and emphasis, abbreviation, punctuation, and spacing, and the measures taken to ensure the accuracy of the transcriptions and annotations included in the edition. The third section briefly outlines the theoretical underpinnings of the linguistic annotation of normalized word forms, word classes and foreign lexis, which characterizes the present edition as a corpus-linguistic one.

## 10.1 Conceptual structure of the edition

The current edition takes a rather ambitious view of the text in ontological terms. Not only does it aim at the representation of both the text and what Genette (1997) labeled as the *peritext*—including things like layout, colour, changes in hand and script, etc.—but it also aims at the simultaneous and interrelated representation of three separate ontologies: the text as a string of characters<sup>3</sup>, the text as an ordered hierarchy of content objects<sup>4</sup>, and the document as a relative spatial arrangement of layout objects. This plurality of ontologies is intended to reflect the observation that no single ontology has been defined that could account for all aspects of the medieval manuscript text, and that different research questions require different views of the text. Furthermore, the juxtaposition and interrelation of these three ontologies can in itself help reveal interesting functional and pragmatic relationships between textual and documentary features.

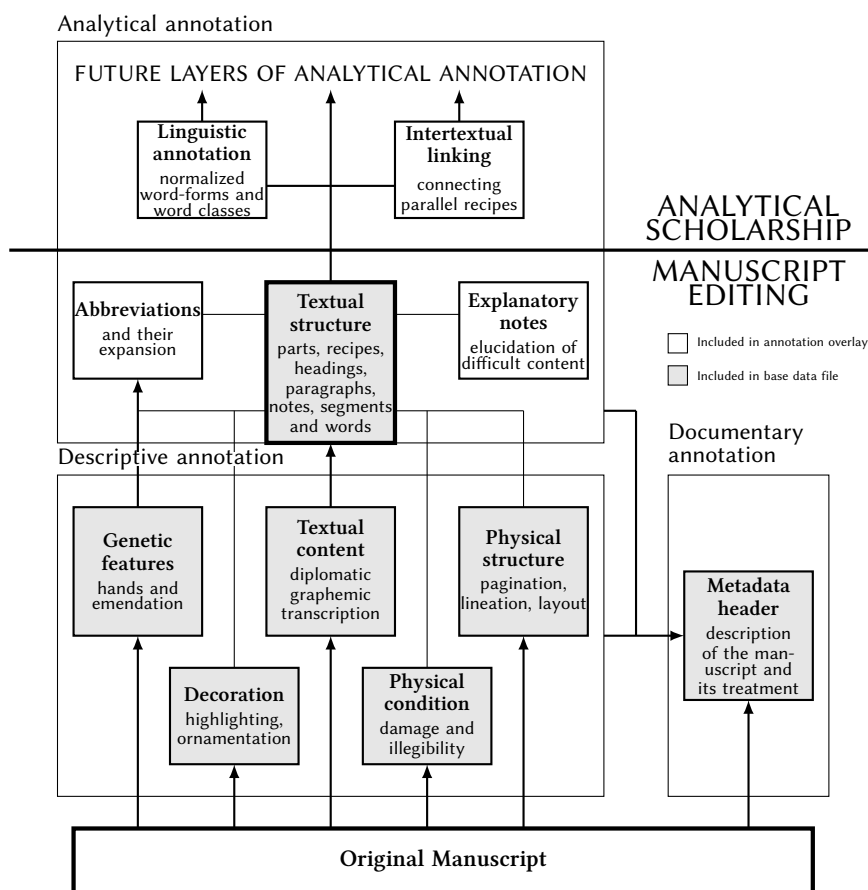
In accordance with this layered approach to annotation, the data and meta-data that constitute the present edition as a *data archive* are divided into three distinct layers: *documentary*, *descriptive*, and *analytical*. On a more practical level, the annotated MS transcriptions contain data relating to a number of different types of textual, palaeographical and codicological phenomena, as well as meta-data describing both the original document and the principles and practices followed in modelling them. In addition to the threefold conceptual division, these types of data and metadata can be categorized according to two further divisions: the practical division of labour between the roles of the editor and the linguist or textual scholar, and the technical division between data encoded directly into the transcription and data encoded as separate annotation overlays. As can be seen in *Figure 10.1*, these divisions—although interrelated—do not correspond exactly with each other: the different types of analytical annotation are divided between the roles of the manuscript editor and the analytical scholar (in this case a linguist or a textual scholar), while the logical structure of the text is the only type of analytical annotation that is integrated with the base data file, other kinds of analytical annotation being encoded in separate annotation overlays linked to this structure. This section will briefly outline each of the three annotation layers, their component parts and their relationships to each other.

### 10.1.1 Documentary annotation

The documentary annotation included for each manuscript included in the present edition consists of a bibliographical identification and description of the edited text itself, including information on its authorship and publication, description of the original source document, and a succinct description of the editorial principles and annotation practices employed in the edition, followed by a classificatory

<sup>3</sup> However, the inclusion of this view of the text does not imply that the text in its entirety forms a single linear string, but rather that it is possible for the user to construct one or more such strings from the textual content of the document encoded in the edition.

<sup>4</sup> In addition to the traditional OHCO, consisting of structural objects like chapters, headings, paragraphs, etc., the edition actually also encodes a parallel, practically but not ontologically integrated hierarchy of *visual formatting objects*, consisting of segments of text visually highlighted in some way in relation to the surrounding (or containing) textual context.



**Figure 10.1:** *The conceptual structure of the edition, indicating the component parts of the three annotation layers, their distribution between the base data file and separate annotation overlays, and the distribution of labour between editorial and analytical scholarship.*

description of the edited document from a corpus-linguistic point of view and a revision history for the edited document. This metadata is presented as structured data within a *TEI Header*, whose general structure is defined in the *TEI Guidelines* (TEI Consortium 2014), and whose application in the present edition is described in section 11.1. The most prominent part of this metadata is the description of the original source document, which describes each edited manuscript in terms of its bibliographical identity (holding repository, shelfmark, title, etc.), its intellectual content (a descriptive list of each separate work or part of work contained in the MS), its codicological properties (material support, collation, layout, scribal hands, decoration, later additions and binding), and its provenance and history.

### 10.1.2 Descriptive annotation

Scholars are increasingly aware of the importance in a manuscript of not only the letterforms made by the scribe but their disposition

upon the page: the use of color, as emphatic or structural or decorative device; the layout of scribal signs upon the page; a hierarchy of scripts within the inscribed text; indications of correction, annotation or deletion; the physical characteristics of the manuscript itself.  
(Robinson 1998: 251)

While both traditional printed editions and linguistic corpora have focused almost exclusively on the ‘text itself’ and its logical structure, a manuscript text always contains more than merely “words plus structure”. Since the satisfactory expression of a manuscript source thus requires a record of more than the ‘text itself’, the majority of the annotation included in the edition falls into the category of *descriptive* annotation, whose aim is to provide a relatively objective description of both the textual and non-textual features of the manuscript text without offering any analytical interpretations on their meaning.<sup>5</sup> Although Renear (2004) has argued that annotating the visual appearance of an element in the source document instead of its function would make the annotation procedural instead of descriptive, this is in fact a misunderstanding, since as (Pierazzo and Stokes 2010: 400) have pointed out, visual description of the source is crucially different from the visual formatting of the output, being clearly descriptive and not procedural.

This descriptive layer of annotation describes a wide variety of textual and paratextual phenomena occurring in the original document and its division into five separate components in *Figure 10.1* is relatively arbitrary. The core element of the descriptive annotation of a document from a linguistic point of view is its *textual content*. This element represents the transcription of what has traditionally been called the ‘text itself’ according to the principles outlined below in section 10.2. However, it should be noted that from a technical standpoint even this element of the edition cannot be entirely expressed using just digital character encoding, traditionally referred to as ‘plain text’, but requires special annotation representing symbols and characters for which even the latest version of the Unicode standard (6.1) with its 110,181 different characters does not yet provide an encoding.<sup>6</sup>

The element of descriptive annotation that is perhaps most intimately connected with the transcription of the textual content is the description of its ‘genetic features’, or of the writing process by which it was produced and of the resulting ontological status of its component parts. In practice this involves the annotation of the scribal hand in which each span of text was written, and of all emendations (additions, deletions and substitutions) made to the text by various agents, including but not limited to the original scribe(s). This annotation is crucial for the purposes of corpus linguistic study, as it associates each span of text to its appropriate animator, as per the requirements outlined by Claridge (2008: 254), and also provides evidence about the emendator’s *stance* to the textual content

<sup>5</sup> It should, however, be kept in mind that purely objective description is an oxymoron, and things like the assessment of the relative degree and cause of damage or an evaluation of the amount of missing text must by necessity remain subjective.

<sup>6</sup> Although Unicode and ASCII translations for most of these special characters are provided in their descriptions within the header (see subsection 11.1.3), this means that a plain text representation (ASCII or even Unicode) of the textual content of the edition will always be merely a rough approximation.

on the most elementary level (i.e. whether it is considered correct or incorrect).<sup>7</sup> On the visual level, the textual content of the document is qualified by annotation describing any *decoration*, applied to it. Decoration is here understood as any visually distinct element in or around the textual content, regardless of whether it is intended to function as a text-organising device or aesthetic ornamentation.<sup>8</sup> Just like for the textual content itself, also decoration—whether a part of the original writing process (e.g. enlarged letters or more formal script) or added after it (e.g. rubrication or underlining)—has been annotated for its genesis by indicating the hand in which it was added, in many cases based on the colour of ink or pigment used.

While the remaining two elements of descriptive annotation—*physical condition* and *physical structure*—are related to the properties of the document as a physical artefact, they nevertheless influence the interpretation of the textual content, either by indicating potential loss of textual content or involving “semiotic codes which readers will decipher—more or less fully, and whether consciously or not” (McGann 1991: 113). Of the physical structure of the document, the edition records its division into pages (folio division being implied by it), which are further characterized in terms of their layout by dividing them into columns and lines and separately indicating the relative position of any elements occurring in the margins. The physical condition of the manuscript, which not only influences the reliability textual readings affected by the damage but can also provide clues about the history of the document, is annotated by recording all physical defects apparent in the document and their implications (loss or illegibility) to the textual content.

However, since the focus of the present edition—and thus of the editorial principles and annotation practices outlined in this and the following chapter—is linguistic, the physical structure of the manuscript is not annotated using the framework for transcribing and encoding manuscript texts from a primarily documentary perspective as physical *zones* and *surfaces*, outlined by Pierazzo and Stokes (2010: 2010) and subsequently integrated into the *TEI Guidelines* as an alternative to the more text-centric default structuring. Instead, the pagination and lineation of the document is expressed by indicating line and page breaks as points in the textual flow (see section 11.3), and any deviations from the default text flow and layout are indicated by transcribing the displaced segment at its logical place in the text flow and separately annotating its position in relation to its place in the text flow and to the physical page.<sup>9</sup> Similarly, also physical damage is annotated in a way that focuses primarily on its effects on the textual content, expressing its

<sup>7</sup> Furthermore, the annotation of all emendations and their assignment to different hands also allows the production of multiple editorial outputs representing the different genetic states of the original document and the examination of the revision or correction process of the text, as suggested by Pierazzo and Stokes (2010: 418).

<sup>8</sup> Indeed, as for example Scott (1989: 31) has noted, the two functions are often inseparably intertwined in many forms of medieval manuscript decoration.

<sup>9</sup> This explicit annotation of layout in relation to the textual logic answers Carroll’s call for preserving the original spatial ordering of the textual elements of recipes (heading, numbering, recipe body, etc.) in editions in order to avoid misrepresenting the manuscript text while simultaneously explicating the logical structure of the text by for example transcribing the number label of a recipe before its title regardless of their positions on the page.

extent in terms of the textual content it affects (see section 11.6).<sup>10</sup>

### 10.1.3 Analytical annotation

Despite the highly diplomatic nature of the present edition and the emphasis placed on the ontological primacy of descriptive annotation of the document, the fundamentally textual nature of corpus-linguistic analysis means that as a corpus-linguistic edition, the most logical choice for the main organising principle of the present edition is logical structure of the document's textual content. This means that instead of the documentary structure of pages and lines, the primary structuring principle of the edition is a layer of analytical annotation that makes explicit the abstract textual structure of the document as an ordered hierarchy of content objects consisting of not only a sequence of recipes with headings and annotations, but also of subgroups of recipes with their own headings, lists of contents and other specialized textual structures, down to the micro-level of individual words, numbers and punctuation characters.

While this annotation reflects the abstract textual structure of the edited text and is thus essentially analytical, it is anchored to the documentary level of the textual object through the limitation of annotated elements to ones that are either implicitly or explicitly signalled by the documentary layer. This means that while recipes are divided into a *front* and a *body*, the former consisting of a heading and possible markers or labels and the latter of one or more paragraphs of running text, for example the internal discourse structure of recipes (ingredients, procedure, closing formula, etc.) is not annotated but is rather considered a topic for a further layer of analytical annotation.<sup>11</sup> The organisation of the data model of the edition around an analytical annotation layer means that this layer must be both integrated with the transcription and prepared by the editor, and neither the division of labour between the editor and the analytical scholar nor the division between integrated and stand-off annotation exactly corresponds with the division between descriptive and analytical annotation, as seen in *Figure 10.1*.

The lowest level of this text-structural annotation, consisting of a tessellated structure of *word-units* integrated into the base data file also serves as a *textual coordinate system* (discussed in more detail below) onto which the other layers of analytical annotation are attached by the means of unique identifiers provided for every such word unit.<sup>12</sup> In addition to textual structure, the other type of annotation included in the base data file that could be considered analytical—although arguably only to the same degree as the transcription of the text itself—is the identification of abbreviations. The editorial *expansion* of the abbreviations, which can be argued to be the properly interpretive or analytical part in the annotation of abbreviations, is also considered an editorial responsibility, but is annotated as

<sup>10</sup> While damage that does not affect the textual content is recorded as potential evidence of the use of the document, its extent and location are given only approximately.

<sup>11</sup> The specific types of textual elements that are used for the annotation are listed and described in section 11.2.

<sup>12</sup> A similar method of annotating each word in the transcription with a unique identifier and using these identifiers as anchors for linking stand-off annotation has also been used by James Cummings in his electronic edition of *The Conversion of Saint Paul* (described in Cummings 2009), although he uses <seg> elements to encode each orthographical word instead of using the <w> element to encode lexical words.

a separate annotation overlay included with the edition, as shown in 10.1. The third type of analytic annotation that is considered to belong to the purview of the manuscript editor consists of *explanatory notes* which are intended to provide the reader with information relevant for the understanding and interpretation of specific contextually bound, highly technical or otherwise obscure passages in the text. Since these notes relate to the content of the textual object as a *work*, they are frequently linked to several parallel versions of the same passage in different MSS, and are therefore stored in a separate document apart from the six transcription documents making up the core data archive of the present edition.<sup>13</sup>

As was observed in section 5.5, the fact that there is neither a technical nor a theoretical limit to the amount of analytical data that could be annotated in a digital edition, the threat of ‘over-annotation’ is especially acute in the context of analytic annotation. While the provision of all the elements of descriptive and documentary annotation listed above is here considered to be the self-evident duty of a documentary editor preparing a digital edition of any kind of historical document, the responsibility for the analytical annotation of the edition is here seen to be shared by the editor and its scholarly audience, based on the principle of “expertise” mentioned in section 5.5. While the editor—as the leading expert of the particular textual object he is editing—is considered to be the most appropriate person for producing the types of analytical annotation listed above (i.e. analysis of textual structure, expansion of abbreviations and elucidation of obscure content), they are also considered to define the limits of editorial responsibility for a digital documentary edition.

Any further layers of analytical annotation are thus considered ‘non-editorial’ in the sense that they are not produced as a part of the editorial process, but as entirely separate endeavours based on different kinds of expertise and motivated either by the specific function of the edition or by the nature of the edited texts. In the case of the present edition, these further layers of analytical annotation include the linguistic annotation of the textual content with normalized word-forms and rudimentary word class information, and the annotation of intertextual links between *parallel versions* of recipes found in the six manuscript versions as defined in section 9.1, the former motivated by the corpus-linguistic orientation of the edition and the latter by the nature of the edited material as a discourse colony surviving in multiple variously related manuscript versions.<sup>14</sup>

### **Word-units as textual coordinates**

Now there is perhaps no unit over which there is less agreement than the word. If there is any agreement at all, it is that the word has to be differently defined for each language analysed. (Bazell 1966: 35)

From a of corpus linguistic point of view, the most natural and convenient unit for a *textual coordinate system* is the *word*, the basic unit of corpus linguistic

<sup>13</sup> The separate documents that contain all of the data and metadata for the present edition are included and described in appendix A.

<sup>14</sup> The possibilities opened up by these two ‘extra’ layers of annotation are exemplified by the analyses undertaken in chapters 12 and 13.

analysis. It is the unit on the level of which most quantitative corpus methods operate, and thus the level on which parts of the text frequently need to be identified and quantified. While units below the word-level—morphemes or individual letterforms—are of interest for many kinds of linguistic research, their annotation would be either considerably more time-consuming (morphemes) or redundant (individual characters).<sup>15</sup> The problem with this approach is that there is no simple uncontroversial definition for ‘a word’ (see e.g. Jespersen 1924: 92, Bazell 1966: 35, Langacker 1972: 37, Matthews 1991: 208, Haspelmath 2011: 32). While the most natural and neutral starting point in written documents would seem to be the orthographic word, the non-standard orthography of Middle English documents makes it impossible to rely solely on orthography for defining word-units in a manner useful for corpus linguistic analysis: Denholm-Young (1954) observes that there “is much faulty word division in minuscule scripts”, and in vernacular manuscripts compounds like *shalbe* occur until the 17<sup>th</sup> century, and the opposite error—splitting what should be joined, as in cases like *be fall*, *ofte sithes*—is also common in the 15<sup>th</sup> century (70). For this reason some exceptions have been made in order to preserve on the one hand the *lexical integrity* (i.e. the ability to occur independently) and on the other the *atomicity* (i.e. non-decomposability) of these annotated word units. In practice this means that:

- a) affixes (like the *y-* or *i-* prefix of the past participle) or other parts of words written separately in the original are joined to their stem word, e.g. *y nogh*, *y made*, *y lyche* and *a boue*,
- b) compound words that are written separately in the original but have an established semantic meaning as a compound that is different from that of their component parts (which might or might not be viable words themselves), are considered to be single compound words, e.g. *there to*, *per of*, *to gyder* and *stoke fysche*,
- c) compound words that are regularly written both separately and together in the original have been consistently annotated as either one or two words, based on Present Day English norms, e.g. all of *anoper*, *an oper* and *a nofer* have been annotated as a single word to maintain comparability of the different forms, and
- d) indefinite articles run into their headword have been separated from it and annotated as separate words e.g. *adele*, *aloffe* and *afeire*.

In making decisions about whether to annotate something as a compound word or two separate words, based on the aforementioned principles, the *Oxford English Dictionary* (OED Online) has been used as a reference.<sup>16</sup> In most cases,

<sup>15</sup> For specific uses requiring access to individual characters within a word, an ad-hoc implementation of the *xpointer()* scheme (DeRose, Maler and Daniel 2002) based either on the use of XSLT functions or on custom software can be used. Although the same system could technically also be used on the level of longer spans of text like the paragraph, it would be neither convenient, efficient nor very robust, as it relies on character positions in a context (the XML document) which not only contains *mixed content* but does not by default treat whitespace as significant (see e.g. Cayless and Soroka (2010)). For research questions requiring reference to individual morphemes within the word-level, the word-units provide a starting point for the morphological analysis of the text, the results of which can be incorporated into the text as a new annotation overlay.

<sup>16</sup> This does not mean, however, that all compound nouns that are listed in the *OED* are annotated as single words; if the meaning of the compound is not essentially different from that of the two words understood separately and both of the components are viable words in themselves—as in the case



the ambiguous nature of these constructions is also signalled by the fact that they occur in the text written both together and apart. It should be noted that there has been no normalization of the original word-spacing—no spaces have been added between words that were written without them or removed from between ones that were written apart, and any hyphens included at line end have been transcribed as they occur in the document. Rather, each word unit has been explicitly annotated as such by enclosing it within an XML element in the manner described in subsection 11.2.5.

In addition to words, also numbers (including any punctuation characters used to identify them as numbers), punctuation characters and independently occurring characters or symbols have been similarly annotated and provided with unique identifiers, creating a *textual coordinate system* which covers all of the transcribed textual content in a tessellated fashion and is used as the primary frame of reference for linking related data and metadata. This means that annotation contained in external annotation overlays can be conveniently linked to individual words or sequences of words of the text using simple *fragment identifiers*. These external annotations can then be used to augment or replace the internal annotation of the edition by a suitable application. The definition of a full-fledged standard format for external annotation overlays and an implementation of that format as a part of a user interface for the edition—allowing the creation and application of such overlays—is beyond the scope of this thesis, although a preliminary sketch for such a standard is presented in the form of the selected analytical overlays included in the present edition and described in sections 11.7, 11.8 and 11.9.<sup>17</sup>

#### 10.1.4 Avoiding overlapping hierarchies

Decades of text-encoding have shown that as soon as we try to mix texts and documents we encounter overlapping hierarchies: textual boundaries do not coincide with documentary ones, apart from some very specific cases, and to handle both at the same time is not possible, especially with a strictly hierarchical markup language like XML.  
(Pierazzo and Stokes 2010: 421)

As was pointed out in subsection 5.6.4, the integration of several ontologically separate annotation layers within a single document poses problems for any structural markup scheme and represents a central challenge for this kind of a layered edition, restricting the choice of data models and annotation structures. As mentioned in subsection 5.7.1, the present edition uses a combination of two different methods for avoiding the problems caused by multiple overlapping hierarchies, one used to reconcile the physical and textual structures of the original document within the base data file and another for allowing the addition of an indefinite number of additional analytical annotation layers to the base data file.

---

of food item names like “green sauce”—the item has been annotated as two separate words.

<sup>17</sup> The development of an online interface for the edition, which will include not only flexible display and corpus search functions, but also the ability to create further annotation overlays and use them for searching and displaying the contents of the edition, is being planned by the present author as a separate post-doctoral project.

### Reconciling textual and physical structure

The corpus-linguistic emphasis and the ensuing primacy of textual structure means that the primary structuring of the textual content of the recipe collections into an OHCO is based on their division into sub-collections, individual recipes, recipe headings and labels, paragraphs of running text, and finally into individual words, numbers and punctuation characters, represented by a hierarchy of nested XML elements.<sup>18</sup> Since these divisions frequently cut across the physical divisions constituted by the pages, columns and lines of the document, these latter units cannot be represented by enclosing XML elements, but are instead indicated by so-called ‘milestone elements’—empty XML elements marking a point in the text—placed at the beginning of each new page, column and line. While this solution does prevent the automatic validation of the latter structures and make it technically more complicated to use the physical document structure as the basis for analysing or displaying the edited text, it does allow the two structures to be more conveniently analysed in relation to each other and the textual content than the use of two separate and parallel encodings recommended by Pierazzo and Stokes (2010: 422). Furthermore, it also allows for the automatic replacement of the milestone elements with enclosing elements in situations where the physical structure is the more relevant one, e.g. for the extraction and display of the contents of a single manuscript page.<sup>19</sup>

In order to facilitate this conversion of milestones into enclosing elements, the methods defined for element fragmentation and virtual joins by the *TEI Guidelines* (TEI Consortium 2014: 626-9) have been used to break up certain ‘textual container’ elements (see subsection 11.2.6) at the physical boundaries denoted by the milestone elements, so that

- word-level elements are broken by milestones denoting line breaks, and
- all textual container elements are broken by milestones denoting column and page breaks.

Although not be necessary in terms of well-formedness or schema conformance, this approach has the benefit of avoiding semantically awkward hierarchical relationships (such as a page number being contained by a text paragraph that continues from one page to the next) and limiting the number of hierarchical levels on which these milestones occur (and thus the number of levels and elements that need to be automatically fragmented when content is being extracted based on these milestones). This approach also serves to make the annotation of physical structure ‘less subordinate’ to the textual structure and is also semantically more appropriate: in an edition emphasizing the materiality of the text, a ‘word divided on two lines’ is a more appropriate conceptualization than a ‘word containing a line break’.

<sup>18</sup> The annotation of textual structure is described in detail in section 11.2.

<sup>19</sup> Appendix E contains a series of XSLT transformations used for the preparation of the diplomatic presentations of the six recipe collections included in this thesis (appendix B), including a transformation for extracting a single page from the XML source files of the edition as a well-formed and TEI compliant XML fragment.

### Annotation overlays

While the above solution is ideally suited for reconciling the two conflicting structural hierarchies encoded within the base data files, it quickly becomes impractical if we attempt to add multiple new analytical structures on top of them. If we wish to enable the application of an arbitrary number of new analytical structures to the data contained in the edition, the only practical solution is the use of separate *annotation overlays*, described in subsection 5.6.4 above. The basic idea of an annotation overlay is simple: it is simply a separate XML document which contains some kind of metadata in the form of XML elements that are associated with the textual content of one or more base data files using the textual coordinate system described above, the technical implementation of which is described in section 11.2. The different annotation overlays included in the present edition, the metadata encoded in them, and the technical means of linking them to the base data file are described in sections 11.7, 11.8 and 11.9.

## 10.2 Treatment of the text

At one end of the spectrum are transcriptions that may be called strictly diplomatic, in which every feature that may reasonably be reproduced in print are retained. These features include not only spelling and punctuation but also capitalization, word division, and variant letterforms. The layout of the page is also retained, in terms of line division, large initials, and so on. Any abbreviations in the text will not be expanded, and in the strictest diplomatic transcriptions, apparent slips of the pen will remain uncorrected.

(Driscoll 2006: 254)

As was pointed out in chapter 5 and section 5.4, digital editions offer a way for sidestepping many of the textual compromises that are traditionally made in printed editions. However, as was also pointed out, the representation of a manuscript text as a digitally encoded text file always involves the classification of analogue phenomena into discrete categories, “a complex business, with many subtle choices having to be made”, which is “far from the mechanical affair it is sometimes thought to be” (Robinson 1998: 253). In the present context of ‘philological’ or *linguistic* editing this involves recognizing the smallest graphical units on the page to which we can assign an alphabetic or symbolic function (discrete *graphetes*), and to represent them with abstract symbols (termed *graphemes*) representing that function. Since this task of separating what Robinson and Solopova (2006: 4) call “marks considered of likely significance” from the background noise of the manuscript page is often dependent on the transcriber’s or editor’s discretion (Rogos 2010: 80), which is likely to be inferior to that of the original scribe who most likely “knew more about the text and its language than we do” (McCarren 1998: 149), the present edition takes a cautious and consequently inclusive

approach to transcription.<sup>20</sup> This means that all distinguishable markings on the page are recorded either by assigning them with a graphemic value, or if this is not possible, annotating their presence and describing them in terms of their visual appearance. In order to establish the relationship between the original manuscripts and their representations in this edition, this section will describe the decisions made in accomplishing this process of abstraction, both on the linguistic level of words and on the orthographic level of the symbols used to represent them.

### 10.2.1 Delimiting the edited text

The first question in treating the text to be edited is the delimitation of the text, i.e. what parts of the textual content of the document are edited. In the case of manuscripts containing only the recipe collection (MSS H4016, H279, As and D), the most natural answer is “all of it”, but in the case of the other manuscript versions the issue is complicated by the fact that the document also contains text belonging to other, more or less related *works*. The material falling outside the bounds of the recipe collection itself can be divided into two categories: 1) text that belongs to another unrelated work, and 2) text that can be considered to belong either to the *paratext* of the collection or to constitute a separate but closely related ‘adjunct’ or ‘ancillary’ text. In the case of the six manuscripts contained here, textual content falling into the first category is found in MSS Ad and C which are miscellany manuscripts containing a selection of different works, while the second category is represented by the tables of contents (a classic example of paratext) and bills of fare prepended or appended to all of the *PD* manuscripts except for MS D.

The approach taken in the present edition is to include not only the recipe collection itself, but also the paratextual and ancillary texts associated with it. This means that in the case of MSS H4016, H279, As and D all of the text contained in the original manuscript is transcribed in the edition, including all later notes added onto their pages, while the transcriptions of MSS Ad and C include all of the text contained on those folia of the manuscript that hold some part of the recipe collection or its immediately adjoining ancillary texts.<sup>21</sup> The decision to include the paratextual and ancillary elements in the edition was based, first of all, on their close codicological integration to the text of the collection; they are not only included at the ends and beginnings of quires containing the recipe collection itself but also frequently end or begin on the same folio as the first or last recipe of the collection (see e.g. MS Ad f. 25r, MS As f. 47v, MS D f. 5r and MS H4016 f. 2r), the most notable exception being MS H279, where both the (incomplete) table of con-

<sup>20</sup> In addition to the avoidance of editorial hubris, there are also two other reasons for this relatively inclusive approach to the textual and paratextual content of the document. First of all, the unavailability of high-quality facsimile images—much less ones that could be made publicly available with the edition—means that the annotated transcription will be the user’s only source of information about the document and should therefore include as much detail as possible. Secondly, even if facsimiles were available, the corpus linguistic orientation of the edition prioritizes analytically encoded and searchable data over the purely graphic representation of a facsimile which cannot be searched or quantified.

<sup>21</sup> The delimitation of the edited text is the most contentious in the case of MS Cosin V.iii.11, where the text of the *Potage Dyvers* version is immediately followed by another, incomplete recipe collection (see appendix F). The boundary between these collections is further blurred by the fact that the quire containing the second collection is missing its outermost bifolium, containing the beginning and end of the collection along with any possible rubrics indicating its beginning.

tents and the bills of fare are separated from the main collection by one or more empty pages. The second reason for the inclusion of not only the clearly paratextual tables of contents but also the accompanying bills of fare was the additional information they provide on the context in which the dishes described by the otherwise relatively uncontextualized recipes were consumed (Weiss Adamson 2004: xviii).

### Inclusion of original annotation

Marginalia and other annotations added to manuscripts by later correctors, readers and other users are usually not included in traditional printed editions, although they may sometimes be recorded by careful cataloguers (Toon 1991: 84).<sup>22</sup> The same applies to many digital editions, including the otherwise very detailed *Canterbury Tales Project* (Robinson and Solopova 2006: 14), even though their inclusion is no longer hindered by the practical restrictions of the printed page. This is an understandable solution in editing printed works, where marginal or interlineal notations made by a reader are usually not seen to “intrude into the life of the text” which has been permanently fixed by the act of printing (Toon 1991: 76), but in the context of medieval manuscripts, the nature of manuscript transmission meant that in annotating the text, making corrections or adding material, the medieval reader or scribe “actually modified the text itself”, as their annotations “became part of the text and were regularly incorporated into the transmission of the text” in the next generation of copying (Toon 1991: 76).

For this reason and in accordance with its meticulously documentary approach, the present edition transcribes also peripheral annotations, whether added by the original scribe or a later user of the text, and whether connected to the original text or not. While many of the annotations are not directly related to the content of the original text, they can potentially offer clues of “what users of texts were up to” when they annotated the text, i.e. what were they doing with the text that led them to leave their own mark in the text (Toon 1991: 75), and thus provide information—however scant—about not only the annotator but also about the cultural and social context in which he or she was using the text; “learning about him, we learn about his period, in what still seems to me the most reliable way – that is, through identification with one member of it” (Jackson 2004: 77).

While the majority of the annotations accompanying the *PD* texts are either trivial, unintelligible, or both, the fact that marginal notes as a general category “are potentially a gold mine” for “many areas in the new scholarly realm of textual studies—for example, reception history, reader response, and the history of reading practices” (Jackson 2004: 78) means that even the most seemingly trivial annotations should be included as a point of principle, since it is not always possible for the editor to know beforehand which annotations will prove to be interesting. In order to allow these peripheral notes to be selectively included in or excluded from analyses or editorial representations, they have been explicitly annotated as such and furthermore classified into *labels* and *markers* referring to the original text in some way, and *detached notes* that are entirely unconnected with it (see subsection 11.4.3).

<sup>22</sup> One such cataloguer is Dr. Ian Doyle who wrote the superbly detailed and extremely useful catalogue description of MS C.

### 10.2.2 Transcription of letterforms

There has never been a single standard convention for the transcription of manuscript texts, and it is not likely that there ever will be one, given the great variety of textual complications that manuscripts—from all times and places—can present.

(Vander Meulen and Tanselle 1999: 201)

Of the transcriptional approaches listed by Robinson and Solopova (2006: 2) and quoted in chapter 5, the present edition—like many other digital editorial projects such as the *Canterbury Tales Project (CTP)* and the WAB (Pichler 1995a: 692)—adopts the *graphemic* principle of transcription, largely for the same reasons as those listed by Robinson and Solopova and discussed in subsection 5.4.1. This means that the principle behind the treatment of orthography in this edition is simple: all words and letters are encoded as they occur in the manuscript, not distinguishing between specific letterforms. The practice, however, is more complex and involves answering two questions: 1) when are two graphically distinct letterforms (*graphetes* or *glyphs*) considered to represent two different letters (*graphemes* or *characters*), and 2) what is considered a word? The answer given to the first question in this edition involves the concept of *allography*, and is discussed below. The second question, on the other hand, is something of a trick question: this edition does not assume the orthographic word (separated by inter-word spaces) to correspond with the morphosyntactic word (as defined in subsection 10.1.3 above), which is annotated separately (see subsection 11.2.5).<sup>23</sup> Thus, in terms of transcription, the scribal word-division—which can provide valuable information about contemporary perceptions of syntactic and prosodic structures and relationships (Lass 2004a: 35)—is represented as it occurs in the manuscript, following the principles set out in subsection 10.2.6 below for the treatment of inter-word space.

As each morphosyntactic word is also annotated with its normalized spelling (see subsections 10.3.1 and 11.9.1), there is no reason to normalize the original orthography in the interest of facilitating corpus searches. This principle also extends to words containing obvious misspellings—quite frequent in culinary recipes, as Hieatt (1998a: 135) has also observed—which are not corrected.<sup>24</sup> Although

<sup>23</sup> This approach is similar to that adopted by the *MENOTA* project which also follows the orthography of the original document in the transcription while explicitly annotating morphosyntactic words (Haugen 2004: 84-5). In addition to the obvious benefit of simultaneously preserving the original scribal orthography and facilitating searching and quantification of the text, Haugen (2004: 85) presents two practical benefits to this approach: first of all, it eliminates any special markup for indicating the omission or addition of editorial whitespace between words, and secondly, the XML element used to annotate the word “offers a convenient location for lexicographical and grammatical attributes, which typically refer to the whole word, not to any constituent of it”, although he does not explicitly mention the use of these elements as a more general coordinate system for anchoring stand-off annotation.

<sup>24</sup> This approach is similar to that adopted by the *WWP* for printed texts and is based on a similar rationale, namely that whatever the source of errors in the text of the manuscript document, they are a part of the text as received by its readers when it was originally circulated, and thus of potential interest to researchers (Flanders 2006: 142). However, unlike the *WWP*, the present edition does not explicitly annotate ‘errors’, since they are a much less clear-cut phenomenon in Middle English manuscript than in modern printed texts, and drawing a line between errors and merely idiosyncratic spellings is practically impossible.

the annotation system would allow for the preservation of both the original and the editorially corrected reading (see TEI Consortium 2014: 79–81), the current approach was selected for two reasons. First of all, although the word intended by the scribe is in many cases obvious, the intended spelling is not, which means that the correction would actually be as much an editorial normalization as a correction. Secondly, a standardized and thus corrected form is already provided for all of the identified words in the original manuscripts, and the correction would thus be redundant.

As Robinson (1998); Robinson and Solopova (2006) point out, it is “not possible to achieve a stringent conformance to any one level [of transcription] in the course of a long transcription of many manuscripts” (Robinson 1998: 253), and consequently most editions do not fall neatly into one of the four categories (*graphic*, *graphetic*, *graphemic* or *normalized*) but incorporate features from several of them. While this edition makes an effort to maintain a consistently graphemic approach, any deviations from it are made rather toward the graphetic than the regularising approach (as described for example in subsection 10.2.4). This approach was chosen after the graphetic approach was deemed impossible to apply with any degree of consistency, based on an examination of the current material in light of the observations made by Robinson (1998 and 2006), Robinson and Solopova (2006), and Rogos (2010) in the context of the *Canterbury Tales Project (CTP)* and discussed in subsection 5.4.1 in chapter 5. The basic principle adopted here is thus the similar to the one outlined by Robinson and Solopova (2006), namely to distinguish between only those and all those “signs held to have potentially graphemically distinctive value” (6).

The concept of ‘graphemically distinctive value’, however, is in itself slightly problematic, being to a large extent an arbitrary matter of convention. Since neither Present Day English nor Middle English have a one-to-one correspondence between graphemes and phonemes, phonemic value cannot be used to define graphemes, and the lack of established spelling system in Middle English means that a semantic criterion is equally problematic. This is the crux of the graphemic/graphetic distinction: how are we to decide which graphetes are merely *allographs* of the same grapheme and which constitute independent graphemes? Graphemes and their relationship to their constituent allographs have traditionally been seen from two different ontological points of view: 1) in terms of family resemblance, i.e. referring to all allographs that are considered to represent the same grapheme, and 2) in terms of systemic oppositions, i.e. whether two forms result in a semantic distinction (Smith 2008: 213). In essence, the first of these views thus sees the grapheme from the point of view of unique and individual *instances* of graphemes, while the second sees it from the point of view of the abstract *system* or typology of graphemes.

Unfortunately, neither of these viewpoints is able to provide a historically stable criterion for dividing the available inventory of graphetes into distinct graphemes. Of the two views, the first is in fact merely a codification of the pre-existing culturally and historically bound perception of the internal relationships within the available inventory of graphetes, while the second is problematic in terms of the highly variable orthography of Middle English: since most words have a variety of acceptable spellings, the majority of Middle English vowels would in many contexts constitute allographic variants of each other rather than separate graph-

emes, as orthographic variation between such semantically and lexically identical pairs as *lytel/lytil* and *wilt/wolt* is commonplace in Late Middle English texts, including the *PD* versions edited here. The fact that even different graphemes can occasionally be substituted for each other with no semantic change would require us to formulate the definition in more absolute terms, defining the allographic forms of a grapheme as those graphemes whose replacement with one another can *never* result in a semantic difference. The problem with this definition, as with all absolute negatives, is that it cannot be conclusively shown to be true, and the fact that a large portion of Middle English writing has not survived means that the likelihood of a semantically significant distinction between a given pair of graphemes existing in some context is non-trivial. A further complication is added if we also consider the diachronic dimension and take as our aim a transcription system that would be suitable for transcribing texts from different historical periods.

### The graphemic inventory used for transcription

Since there are no truly objective and ahistorical criteria available, the decision must needs be an arbitrary one. However, as Wyatt (1987: 167) has argued, the arbitrariness of editorial decisions is not necessarily a problem, if they are consistently adhered to. For this reason, the approach taken in this edition has been chosen principally on practical grounds, being simply the one considered most convenient for transcribing historical documents for the purposes of corpus linguistics and for allowing the preservation of transcriptional consistency over different types of texts. On these grounds, the graphemic inventory of alphabetical characters to which the transcriptions of the present edition are normalized consists of the modern alphabet supplemented by the Middle English characters *yogh* (ȝ) and *thorn* (þ). While it could be claimed that this selection is not only arbitrary but also anachronistic, there are also other justifications besides convenience.<sup>25</sup> The other possible option for a graphemic standard would be to adopt the Late Middle English practice. There are several problems with this. First of all, it is not much of a standard. The usage of letterforms by scribes varies: a distinction which for one scribe is clearly graphemic might be purely graphetic for another. Secondly, the Middle English practice is less distinguishing than later language forms with its graphetic alteration of pairs like ⟨u/v⟩ and ⟨i/j⟩.<sup>26</sup> and the present editor feels more comfortable erring on the side of graphetic distinction than of normalisation.

However, one could also argue that distinguishing between ⟨u/v⟩ and ⟨i/j⟩ would also make it logical to distinguish between at least ⟨f⟩ and ⟨s⟩, since this pair has a similar status in Middle English, being guided by the position of occurrence within a word. For a counterargument, the issue can be viewed from a diachronic perspective; although the distinctions between ⟨u/v⟩ and ⟨i/j⟩ were not semantically or phonetically significant in the Late Middle English period, they

<sup>25</sup> Typographic convenience is here not considered a valid justification, since the introduction of Unicode (The Unicode Consortium 2011) and the Medieval Unicode Font Initiative (Haugen 2009) and tools like FontForge <<http://fontforge.sourceforge.net/>> means that anyone can become their own typographer and design a font suitable for representing any special characters deemed necessary.

<sup>26</sup> The convention of using ⟨i⟩ and ⟨u⟩ for vowels and ⟨j⟩ and ⟨v⟩ for consonants did not gain general acceptance in England until the 18<sup>th</sup> century (Hector 1966: 40).



have since become significant, while the *f/s* distinction has never been significant in the history of English.<sup>27</sup>

The inclusion of ⟨ȝ⟩ and ⟨þ⟩ is based on their interpretation as graphemes in their own right. Although ⟨ȝ⟩ could etymologically be considered a variant glyph of ⟨g⟩, it can be argued to have been phonetically distinct by the Middle English period. The *thorn* on the other hand can be considered an independent grapheme since it is not phonemically equivalent to any other single grapheme, although it can be seen as the equivalent of the bigraph ⟨th⟩; its status could be defined as something between a grapheme and an abbreviation marker, which, incidentally, are also treated as graphemes in the present edition (see subsection 10.2.4 below).

The definition for the graphemic system for the present edition could thus be defined diachronically as consisting of those sets of graphemes that have in some context at some point in the history of English language formed an independent grapheme contrasting either phonemically or semantically with other contemporary graphemes.<sup>28</sup> This solution of adopting the modern graphemic system extended with the Middle English graphemes has the benefit of diachronic inclusiveness, i.e. the same graphemic system can also be justifiably be used for several other periods with the same level of distinction for each period, which means not regularizing anything that cannot be justifiably regularized for all of the periods of the English language.

Since the specific interpretations of what this *graphemic transcription* entails in terms of specific textual phenomena can vary, the following list identifies some specific decisions made for this edition:

- *i/j* and *u/v* are treated as separate graphemes, although they are not semantically contrastive in Middle English;
- dots and ticks over ⟨y⟩ or ⟨i⟩ are treated as a part of the letter and not transcribed;
- all *puncti* and other symbols normally used for punctuation are transcribed even when they are used around numerals or the capital *I* (personal pronoun), the differences in their function being reflected in their annotation as either punctuation characters or parts of a number;
- no distinction is made between the different forms of the tironian *et*, the forms with and without a horizontal cross-stroke being considered the same symbol;
- all flourishes and abbreviation markers are transcribed, regardless of their semantic function.

<sup>27</sup> From a practical point of view it could also be noted that while *f* can quite reasonably be considered subordinate to *s*, it would be difficult to tell which one of *u* and *v* would be the superordinate form to which the other could be normalized, although it is clear that this practical issue is merely a reflection of the historical development of the *who* contrast pairs.

<sup>28</sup> This means that the *wynn* (ƿ), and the *ash* (æ) are also considered independent graphemes and would be represented as such if they occurred in the texts edited here, since they were considered independent graphemes in the Old English period. The case of the *eth* (ð) is slightly ambiguous in terms of this definition, since it seems to have been used completely interchangeably with the *thorn*.

### Ambiguous letterforms

As Pidd and Stubbs (1997) have pointed out, even a graphemic system of transcription which does not differentiate between variant letterforms in the final transcription still involves the *graphic* and *graphetic* levels in the transcriber's interpretative act, as the successful graphemic representation of a manuscript text requires "that the graphetic nature of the individual characters be understood, because recognizing that the many different forms of *e* are all significant of *e*, although not transcribed individually, is fundamental to not mistaking one of its forms, such as the 'round' *e*, as actually being an *o*" (56). As many scholars have noted, most fifteenth-century hands with their ambiguous combinations of minims provide many possibilities for misunderstanding if a word is not recognized correctly based on the context.<sup>29</sup> This ambiguity is exacerbated by the fact that not all scribes "indicate the position, or even the presence, of *i*, and after 1200 very few of them make any visible distinction between *n*, and *u*, *m* and *ni* (*in*, *ui*, *iu*), and so forth" (Hector 1966: 27). In some hands also individual letterforms in the manuscripts can be ambiguous, as is the case with the graph *ȝ* (yogh), which is used for both 'ȝ' and 'z' by many scribes (Lucas 1998: 174), including the ones responsible for most of the surviving *PD* versions.

Due to these ambiguities, there is more than a grain of truth in the paradox that the "legibility in a manuscript document consists chiefly in the reader's prior knowledge of what it contains" (Hector 1966: 14). In the present edition, graphically ambiguous characters have been interpreted and transcribed according to their context, which does in many cases allow for an unambiguous interpretation of the letterform without appreciable uncertainty. However, while culinary recipes as a genre have the benefit of being relatively formulaic and predictable in their lexical and syntactic choices, they also contain a fair amount of originally foreign dish names and other relatively obscure terms that have been corrupted into unrecognisable forms through copying. This means that the transcriber often has to approach words with no more contextual information than that the word represents something edible, resulting in situations where it is impossible to know whether a character in the name of a dish is intended as a ⟨u⟩ or an ⟨n⟩. In transcribing such words, the most likely option, based on comparison with parallel versions and the known lexis of culinary recipes, has been chosen and the uncertainty indicated through a special annotation (see subsection 11.6.2).

### Non-alphabetic symbols

In addition to the selection of alphabetical graphemes described here, the present edition also records a number of other symbols, including a variety of special abbreviation markers (discussed in subsection 10.2.4 below), hyphens, punctuation characters (discussed in subsection 10.2.5 below), and some miscellaneous symbols like various kinds of reference markers (such as asterisks, carets and crosses) used to indicate points of insertion for additions or as marginal markers.

In the present edition, also abbreviation markers consisting of a superscript letter are considered to belong to this category of non-alphabetic special symbols

<sup>29</sup> As an example of this, (McCarren 1998) points out that "paleographically *hi* has the same makeup as *lu*, a vertical stroke and two minims".(150)

for two reasons. First of all, their function is clearly different from their alphabetic value, which only represents a part of the textual content they are used to stand for. Secondly, also their form frequently differs from the corresponding letter when used as a normal part of the script.<sup>30</sup> The non-alphabetic quality of the superscript abbreviation markers is also apparent in the treatment of the ‘2-shaped’ superscript abbreviation for ‘-ur’ which is technically a superscript round ⟨r⟩ but is practically never treated as such but seen as a separate symbol in its own right.<sup>31</sup> The present edition extends the same treatment to all superscript characters that are used to mark an abbreviation. Superscript characters that stand only for themselves, however, are annotated merely as visually distinct versions of the alphabetical character (see subsection 11.5.1).

Since the graphemic identity of these symbols is much less definite and the specific function of many of them is still relatively unknown, their abstraction into graphemes has been treated more conservatively than that of alphabetical characters, erring rather on the side of the graphetic. Due to this indeterminacy, all non-alphabetic symbols that are deemed sufficiently different from each other in terms of their graphical features or their genesis are here treated as independent graphemes, even if their general function would seem to be the same. A complete list of the non-alphabetic symbols encoded in the present edition, complete with their method of annotation and a brief description, is included in subsection 11.5.3.<sup>32</sup>

### 10.2.3 Capitalization and emphasis

Since the pragmatic significance and intentional nature of capitalisation and other types of visual emphasis as a means of text-organizing device has been frequently demonstrated over the last decade (see e.g. Caie 2003, Dollinger 2004, Carroll 2006, and Tyrkkö, Marttila and Suhr 2013), special attention is paid to it in the present edition. Unlike in printed texts with its separate upper and lower case types, the distinction between ‘capital’ or majuscule and ‘regular’ or minuscule letters in manuscript texts is much less clear cut. Robinson and Solopova (2006: 12-3) have approached this distinction by introducing the concept of “emphatic” and “unemphatic” characters to describe the system of capitalization and emphasis in the *Canterbury Tales* manuscripts. They define emphatic characters as a category including “capital letters of various sizes, differently emboldened letters, and ornamental capitals”, often emphasized using colour, which work together with “paragraph marks, layout, and punctuation” to organize the text. The solution of

<sup>30</sup> The most notable example is superscript ⟨a⟩, which is in court hands always written differently than the ⟨a⟩ written on the line, resembling either the ‘cc-form’ of ⟨a⟩ or a *u*, possibly topped with a horizontal stroke (Denholm-Young 1954: 67).

<sup>31</sup> The *CTP* constitutes an exception to this, as they treat both this symbol and the ‘superscript squiggle’ as allographic versions of a superscript ⟨r⟩, based on the fact that both of them are used to represent *-ur*.

<sup>32</sup> A special font is used to represent these in the diplomatic representations of the present edition included in this thesis. This font is based on the *Junicode* font, designed by Peter Baker and available at <<http://junicode.sourceforge.net/>> under the Open Font License (<<http://scripts.sil.org/OFL>>), and has been further expanded with glyphs for a number of abbreviation markers and other special characters not included in the original font using the Open Source *FontForge* software (<<http://fontforge.org/>>). This modified version of the font (*JunicodeDECL-Regular.ttf*) is included with the hypertext reading edition contained in appendix D.

the *Canterbury Tales Project* was to use capital letters (along with some annotation) to indicate emphatic characters, but the present edition takes a slightly more detailed approach to textual emphasis, indicating not only the fact of emphasis but also its type.

Considering the importance of capitalization and the use of *litteræ notabilioris* as a text-organizing device—which has been called “the most significant development for the history of punctuation in the narrowest sense of the term” by Parkes (1992: 34), this twofold distinction was judged to be too simplistic for the purposes of the present edition.<sup>33</sup> Instead, this edition adopts a more complex multidimensional annotation system, where the size, shape, colour and possible decorative ascenders and descenders of emphatic letters are annotated separately. In the present edition, simple capitalization is used to indicate those letters that are set apart from the surrounding text (and the customary minuscule form of the letter in question) by their *form*, being clearly intended as capital forms of the letter. The capital forms of letters are understood to be normally slightly larger than their minuscule counterparts, commonly having a body height equal to the full height of minuscule letters, including any ascenders.

As recommended by Driscoll (2006) for editions concerned with textual structure, separate annotation is used for indicating letters whose size and shape do not match, i.e. *large minuscules* or “letters that have the shape of minuscules but the size of majuscules” (256) which are clearly used for emphasis when no capital or majuscule form of a letter exists in the hand, and *small capitals* or “letters that have the shape of majuscules but the size of minuscules” (256). This creates a two-dimensional fourfold classification of letters, in which the unmarked minuscule is the norm and the three other classes can be considered to be emphatic forms.<sup>34</sup> In addition to these categories, decorative initials of unusual size (the height of two or more lines of text) are annotated separately, as are letters written in super- or subscript.<sup>35</sup>

In addition to the size and shape of characters, the present edition also annotates the use of coloured pigment either for the whole letter or for highlighting by a coloured stroke (a common way of marking the initials of important words within the text). Finally, decorative flourishes added to the ascenders or descenders of letters (and occasionally also the last foot of a letter without a descender), mainly on the first and last lines of the page, are also annotated.

### 10.2.4 Abbreviation

In the later Middle Ages the use of abbreviations is so widespread and often so careless that their extension is a matter of great difficulty. The

<sup>33</sup> Even Robinson and Solopova (2006: 12) admit that the twofold distinction is “a simplification of the system found in the manuscripts, as a result of which some information is lost”.

<sup>34</sup> The form ⟨ff⟩, which was used as the capital form of ⟨f⟩ from the second quarter of the 13<sup>th</sup> century onwards and persisted to the 18<sup>th</sup> century in court hands (Denholm-Young 1954: 33), is transcribed as *ff* instead of *F*, following the graphemic transcription principles described above.

<sup>35</sup> Super- and subscript characters used as abbreviation markers are annotated as specific symbols, separate from characters super- or subscripted merely out of habit or for considerations of space, as they are considered to have a separate identity and function distinct from their letter identity, as is indicated by the highly specific shape of such superscript characters as the flat-topped superscript ⟨a⟩.

‘m or n’ sign is often used in a meaningless way, having degenerated into a mere flourish. These have been the cause of much inconsistency and vacillation in the transcription of manuscripts for editions of English texts. (Denholm-Young 1954: 69)

As the comment of Denholm-Young quoted above implies, the ontological definition and subsequently the editorial treatment of abbreviations has been a contested topic. As Driscoll (2006) points out, the entire term *abbreviation* is often used ambiguously, sometimes being used for “the mark, sign, or letter (if there is any) that indicates that something has been suppressed” (259) and sometimes for the entire word of which a part has been suppressed. While Driscoll suggests a distinction between “abbreviations with a lexical reference (suspensions, contractions, and a number of brevigraphs) and those with a graphemic reference (superscript letters and signs and the remainder of the brevigraphs)” (259), this distinction does not address the underlying ontological ambiguity of the concept.

In order to resolve the ambiguity of the term *abbreviation* and to allow abbreviations to be treated in a consistent manner, the present edition adopts the ontological distinction made by the *TEI Guidelines* (TEI Consortium 2014: 353-6) between *abbreviated words* and *abbreviation markers*, the former referring to words that have been partially suppressed and the latter to any symbol indicating such a suppression. The same distinction naturally applies also to the expansion of abbreviations, the *expanded word* referring to a word form of which a part has been supplied editorially, and the *editorial expansion* to the specific graphemes supplied by the editor in expanding the word. Once we have separate terms for the two distinct levels of the phenomenon, we can leave the general and ambiguous term *abbreviation* to refer simply to the phenomenon or act of using an abbreviation marker to indicate the presence of an abbreviated word. In light of this distinction, Driscoll’s division can be reformulated as one between cases where the abbreviation marker and the abbreviated word coincide in their extent, and those where the abbreviation marker represents only a part of the abbreviated word, or from another point of view, between cases where all of the graphemes of the expanded word have been supplied by the editor and those where some of the graphemes making up the word are present in the document.

### Abbreviation markers

As mentioned above, this edition transcribes not only alphabetical characters, but also all other symbols found in the original manuscript, regardless of their significance or function, including a variety of abbreviation markers. Whereas for example the *Canterbury Tales Project* does not record any abbreviation markers or other characters that are judged to be otiose flourishes with no semantic function (Robinson and Solopova 2006: 4), this edition transcribes all recognizable symbols on the page equally and makes the distinction between abbreviations and otiose flourishes through their annotation.<sup>36</sup> In terms of their identity, the present edi-

<sup>36</sup> The exclusive approach adopted by the *CTP* is in fact extremely risky, as the judgement involved is often extremely difficult to make. As Stokes (2010: 240) quite soberly points out, if even a palaeographer as eminent as Malcolm Parkes (1969: xxix-xxx) was forced to concede that “he cannot resolve the ambiguities between otiose strokes and abbreviation strokes in fifteenth-century cursive”, it seems quite foolhardy to expect a research assistant or a beginning editor to be capable of the task.

tion treats abbreviation markers roughly analogously to alphabetical characters: each ‘type’ of abbreviation marker that is considered to constitute an independent grapheme is transcribed as a separate symbol. The graphemic status of abbreviation markers, however, requires some clarification. Robinson and Solopova (2006: 4–5) mention abbreviations as an exception to their graphemic transcription principles, since they are “categorized, to some extent, by their graphic realization in the manuscripts” instead of purely by their semantic referent (i.e. the expansion they imply).

Although this differentiation on the basis of graphic features does indeed lean towards the side of graphetic transcription, it can be justified by two (related) ways in which the symbols used for abbreviation differ from other kinds of symbols: First of all, the medieval system of abbreviation “did not aspire to the mechanical precision of a shorthand in which each symbol is given a constant equivalence; and the shape of the mark is not always a certain indication of its function in its context” (Hector 1966: 29).<sup>37</sup> This vagueness in the meanings of different abbreviation symbols, combined with the general idiosyncratic tendencies of medieval scribes, means that two symbols might be used identically in one text and quite differently in another. For this reason alone, it would be more prudent to again err on the side of making too fine a distinction than to blithely conflate two different symbols with the same function (and encoding) on the basis of just a few texts. The second, related, point is that there has been too little research on the usage patterns of abbreviations by different scribes, in different kinds of texts, and in different periods for us to make sufficiently informed judgements on the graphemic status of many abbreviation markers.<sup>38</sup>

For these reasons abbreviation markers, although treated similarly to alphabetical characters, have slightly different—more graphically oriented—criteria for being treated as independent graphemes. Due to the greater number of unknowns, all abbreviation symbols that are deemed sufficiently different from each other in terms of their graphical features or their genesis are here treated as independent graphemes, even if their customary function would seem to be the same. This means that many symbols that have on the basis of their function been considered forms of the same symbol by other scholars (see e.g. Hector 1966: 31, 33 and Denholm-Young 1954: 67) are here identified separately. Examples of these kinds of related symbols which could also be considered allograph include the ‘superscript squiggle’ (ʹ) and the ‘superscript 2’ (ʹ),<sup>39</sup> and the ‘hook’ (ʹ) and the ‘tailed hook’ (ʹ).<sup>40</sup> The annotation of abbreviations, their expansions and the various symbols used as abbreviation markers are described in section 11.7 and subsection 11.5.3.

<sup>37</sup> Many abbreviation symbols also vary in significance according to the language of their use context (Denholm-Young 1954: 70), and many abbreviation markers “which are special or significant when they occur in contractions may be nothing of the kind when they terminate suspended forms” (Hector 1966: 29).

<sup>38</sup> This is reflected in the great variety of ways in which abbreviation markers have been categorized and classified in the literature.

<sup>39</sup> The treatment of these two symbols as separate graphemes instead of mere hand-specific allographs is also supported by the observation of Rogos (2010: 84) that both of them are frequently used in the same text by the same scribe.

<sup>40</sup> Hector (1966: 33) also considers the *tail stroke* (ʹ) to be a variant of this same symbol, which he lists as Sign No. 9 in his treatment of abbreviation markers.

It should also be pointed out that unlike Robinson and Solopova (2006: 8–9) and consequently Rogos (2010: 80–1), this edition does not treat letters whose ascenders are crossed by a horizontal stroke (most notably ⟨ñ⟩ and ⟨ll⟩) or that have a diagonal flourish stroke appended to them (like ⟨ṭ⟩ or ⟨ṛ⟩) as unitary letter forms but rather as combinations of a letter and an abbreviation marker (a horizontal superscript tittle or a diagonal tail-stroke) that may either be otiose or signify an abbreviation. As for other abbreviation symbols, the graphemic transcription is the same in either case, the difference in abbreviatory status being indicated through the annotation of their expansion or the lack of it.

### Expansion of abbreviated words

While printed diplomatic editions have traditionally been forced to represent abbreviated words either as they appear in the manuscript ('strictly diplomatic editions') or as expanded by the editor ('interpretative diplomatic editions'), the use of annotation allows digital editions to encode both within the same edition and make use of whichever form is more useful for a specific analytical task.<sup>41</sup> While the expansion of abbreviations serves many useful purposes, an accurate representation of the original abbreviated form in addition to the editorial expansion is especially important for quantitative studies of textual variation, as each abbreviated word represents a separate orthographic variant, and an encoding indicating just the presence of an abbreviation but not its orthographic realisation would run the risk of misrepresenting the extent of orthographic variation and hiding linguistically salient features of the text (Wright 2000a: 154; Johansson 2004: 103).

While the inclusion of also the original abbreviated word relieves some of the pressure for successfully reconstructing the spelling that the abbreviation is intended to represent, the usefulness of and justification for the expansion still rests largely on some resemblance between the editorial expansions and the original scribal orthography. The minimal requirement for editorial expansion of abbreviations could thus be considered to be that they do not violate the scribe's customary spelling habits (Lucas 1998: 173). However, in order to do full justice to the medieval textual *variance* in expanding abbreviations, the editor would need to first discover the rules governing the variation in the scribe's orthography in every detail and then expand each occurrence according to these rules (Wright 2000a: 155). Since this is practically impossible, we must settle for a less ideal solution and accept that the editorially expanded forms can never attain the status of original scribal text, restricting their use to purposes for which their convenience outweighs their orthographic inauthenticity.

As Haugen (2004: 79) notes, the expansion of abbreviated words is not always straightforward, as some abbreviation markers have several possible expansions, making the exact orthographic form that any expanded word should take always conjectural. For a practical solution for expanding abbreviations, a "common and reasonable practice" also adopted in the present edition, is to expand words spelled in multiple ways "into the full form most often found in the text" (Marvin 2004: 25–6), assuming that such a form occurs in the text. However, as both Marvin and Wright (2000a) point out, the fact that medieval languages "did not observe regu-

<sup>41</sup> Dynamic digital presentation techniques allow the concurrent representation of both the abbreviated and expanded forms to be extended also to the presentation layer.

lar spelling conventions” (Wright 2000a: 152) means that this approach does run the risk of artificially overrepresenting this form and imposing “a spurious *uniformity* on a medieval text” (154). For this reason, the editorially expanded word form should always be treated as merely conjectural (Haugen 2004: 79), and the original abbreviated forms used for any linguistic research regarding spelling variation or otherwise dependent on original scribal orthography, with each specific abbreviated form treated as a separate orthographic type.

Two special cases that pose an especially difficult challenge include highly formulaic abbreviations (Denholm-Young 1954: 81) and abbreviated words whose linguistic identity is unclear (Wright 2000a: 152). Examples of the first case in the present edition include closing formulae such as the abbreviation  $\tau \xi . f.$  used in MS H279 to indicate the end of a recipe (expanded as ‘*andserue forthe*’ based on unabbreviated instances elsewhere in the MS) and the abbreviated ordinal numbers of the format *Ca. xv<sup>m</sup>* used in D and expanded partially as ‘*Capitulum xv<sup>m</sup>*’, treating the superscript *m* not as an abbreviation marker but as an ordinal indicator constituting a part of the number, analogous to the superscript *th* used in PDE. Of the latter case, the most obvious example in the present edition is the abbreviation marker  $\tau$ , which is expanded either as *and* or *et*, depending on the surrounding linguistic context.

### Graphemic ambiguity in abbreviated words

As Robinson and Solopova (2006: 10-1) point out, abbreviations also involve transcriptional problems not directly related to the graphemic identity of the abbreviation marker itself. Perhaps the most difficult—and most common—of these is the ambiguity caused by word-final abbreviation markers, especially by a tittle placed over two minims. While this edition avoids some of the problems associated with word-final markers with an uncertain abbreviatory status by transcribing all abbreviation markers and other symbols regardless of their significance, the ambiguous graphemic status of the two minims with a ‘superscript tittle’<sup>42</sup> remains problematic: should they be transcribed as ⟨u⟩ or ⟨n⟩ in the case of words like *heroū/heroñ*, where the tittle could be taken either to signify the omission of ⟨n⟩ (in which case the preceding letter should be interpreted as a ⟨u⟩) or to be an otiose flourish (in which case ⟨n⟩ would be the appropriate interpretation of the preceding letter).<sup>43</sup>

This edition adopts a practice similar to that used by Robinson and Solopova (2006: 10-1) by erring on the side of significance in the treatment of the superscript tittle, treating it as a significant abbreviation marker in all instances where it is at all plausible. Thus for words ending in a combination of ⟨o⟩, two minims and a tittle the preferred interpretation is *-oun*, unless the context suggests otherwise (as in the case of *boū/boñ* with the sense of ‘bone (n.)’, where the latter option has been chosen). This decision is supported not only by the frequent occurrence of unambiguous *-oun* endings in the *Potage Dyvers* manuscripts but also by similar

<sup>42</sup> Robinson and Solopova (2006) call the horizontal superscript symbol a *macron*, but due to the association of this word with the Greek diacritic symbol indicating a long vowel, the term *tittle* (from Latin *titulus*) is used instead.

<sup>43</sup> In some cases even a word-final ⟨e⟩ is an option, in which case the preceding letter would again be ⟨n⟩.



observations by Robinson and Solopova (2006: 10-1).

### 10.2.5 Punctuation

As crucial as an editor's choice of words and textual layout, yet at the same time invisible or silent to most readers from its sheer familiarity, is punctuation. (Cooper 1998: 86)

Although it has often been argued that medieval scribes used punctuation “scantily, inconsistently” (Markus 1997: 214) and “so indiscriminately that it is normally a waste of time to reproduce their efforts in modern transcripts” (Denholm-Young 1954: 77), some scholars (e.g. Marvin 2004) have over the last decade come to acknowledge that while decidedly different from modern practice, medieval punctuation does often provide the reader with guidance in terms of the syntactic and structuring of the text. The very fact that relatively little is still known about the functions of medieval punctuation should be taken as a strong argument for representing it as faithfully as possible in order to provide material for its detailed empirical study.

The treatment of punctuation in the present edition follows the same logic as the one followed for words and their constituent characters. While punctuation has not been modernized, the various constellations of dots and lines used for punctuation in the manuscripts have been represented on a graphemic level as combinations of a limited set of punctuation symbols.<sup>44</sup> Although the core of the late medieval punctuation system—consisting of the *punctus*, the *punctus elevatus*, the *punctus interrogativus* and *litterae notabiliores* (Parkes 1992: 42)—was already quite well established, both the shapes and the functions of punctuation symbols still “varied considerably in the pre-printing era” (Parkes 1992: 2), making their assignment to distinct graphemic categories is much less certain than in the case of the more established alphabet. The graphemic representation of punctuation symbols is made even more ambiguous by the fact that according to Parkes, there was no one-to-one correspondence between the shape of punctuation symbols and their function; “[c]ertain marks with different shapes are graphic variants of the same symbol and share similar functions across the centuries” while some “symbols with similar shapes [...] have different functions at different stages of the history of punctuation” (2).

Although punctuation symbols are here encoded according to their form without regard for their function beyond their role as punctuation, this discrepancy between form and function does pose a problem for the graphemic encoding of punctuation symbols, since graphemic abstraction does to a certain degree depend on the *function* of the item. The practical approach taken in this edition is therefore to assign each symbol identified as a punctuation mark to the grapheme

<sup>44</sup> These symbols, as well as all other ‘special symbols’ or non-alphabetical characters are listed in subsection 11.5.3.

whose known graphetic realizations it most closely resembles without considering its position or function in the text.<sup>45</sup> This means that the specific character values used to encode punctuation characters should not be taken to imply a certain grammatical or rhetorical function, but should be seen as interpretations of the visual appearance of the symbol.<sup>46</sup>

### 10.2.6 Spacing

As was explained above, the transcription itself reflects the spacing of the original document and not the division of the text into word-units, which are annotated separately. While the spacing used to differentiate word-units from each other in manuscripts is a gradated phenomenon, the encoding of white space in a digital form is traditionally a binary one. In other words, while the amount of space between associated sequences of graphemes—i.e. words—on the manuscript page varies from one instance to the next, the digital representation simply does or does not contain a whitespace character. Since it is not practicable (or in some cases even possible) to measure and quantify the amount of space between adjacent graphemes, this edition settles for this traditional binary approach.

The judgement between whether there is or is not a significant space between two adjacent graphemes is, of course, a subjective one, although in many cases the choice is clear: either there is ample space or the graphemes are written without any space between them. In ambiguous cases where the space is larger than is typical between graphemes of the same word but smaller than the usual inter-word space employed by the scribe, the transcription errs towards normalisation: if the space is located within a unit that is traditionally considered a single word, no whitespace character is inserted, but if it is located between what are traditionally considered to be separate words, the space has been transcribed as a whitespace character. Any unusually wide spaces, larger than the width of the grapheme *m* (or three minims) have been transcribed using the <space> element, as described in subsection 11.3.4.

### 10.2.7 Accuracy and proofreading

Since transcription and annotation as human activities are always prone to error, various measures have been taken in the production of this edition to ameliorate the effects of human error. The texts edited were originally transcribed by the editor either from grayscale digital images (MS Harley 4016) or from microfilm copies (digitized by the editor) provided by the holding library. Subsequently, each text was proofread twice, once against the digital surrogate copy and once against the original at the holding institution. The proofreading against the original also involved the annotation of various manuscript features not visible in the grayscale images, such as changes in hand and ink, decoration and highlighting

<sup>45</sup> This means that punctuation symbols used as marginal *notae* are encoded identically to those occurring in a more traditional punctuating function within the textual stream, apart from the fact that they are annotated simply as independent characters instead of punctuation characters (see subsection 11.2.5).

<sup>46</sup> This approach differs from the approach adopted by many otherwise relatively diplomatic transcription systems, such as the *MENOTA* system described by Haugen (2004), which normalize the original manuscript punctuation to conform to modern conventions.

and manuscript damage, as well as the examination of the physical properties of the manuscript required for its description (see section 9.2).

Since the annotation of manuscript features (as well as many textual features such as abbreviation and layout) was originally done in a shorthand markup using word-processor-based tools developed by the editor, the conversion of this shorthand markup to TEI-based XML annotation involved a further round of proofreading and validation, focusing mainly on the optimization and semantic validity of the annotation. As Blake (2000: 34) has observed, the linguistic analysis of an edited text is an efficient way of discovering any remaining transcription errors, which will often show up as aberrant results. This was also the case with the linguistic annotation of the edited texts with a normalized word form and a word class for each word token, which also served as a final round of close proofreading. This final round of proofreading involved the evaluation of each of the roughly 120 000 words in the edition in its context to ascertain its identity and function in the text. This had the additional benefit of bringing up a handful of remaining transcription errors and ensuring that all words in the edition are properly tokenized and annotated according to the definition of a word given in subsection 10.1.3.<sup>47</sup>

### Accuracy, consistency and ‘richness’

As Robinson and Solopova (2006: 14-5) have observed, the concept of *accuracy* in manuscript transcription is intimately tied to the concepts of *consistency* and *richness*, both of which can be seen to have an inverse relationship to accuracy, albeit in different senses. If we take consistency to mean the principle of representing all occurrences of a given phenomenon in the manuscript using the same encoding, the inconsistency inherent in the manuscripts themselves can create situations where the accurate representation of the manuscript will result in symbols and annotations being used inconsistently in the transcription. This edition tries to alleviate this dilemma by separating the description of form from the description of function on several levels, most importantly for abbreviations and for morphosyntactic words. This means that instead of for example having to choose between using the appropriate graphemic representation and implying an inappropriate expansion and indicating the appropriate expansion by using an inappropriate graphemic representation, it is possible to have both an accurate graphemic representation of what appears on the manuscript page and an editorial interpretation of what it most likely means in the given context. In terms of words, the same approach is taken in relation to both spacing and spelling, as described above: the transcription reflects the visual reality of the original manuscript, while the explicit annotation of word-units and normalized word-forms provides the semantically consistent editorial interpretation.

In terms of richness, the relationship is more straightforward: “the more detailed the transcription, the more possibility there is for error” (Robinson and Solopova 2006: 15). In this regard, there is much less that can be done by means of editorial practice, and apart from the safeguards described above, the current edi-

<sup>47</sup> The initial annotation of word-units was done algorithmically based on whitespace with special linking and separation annotations indicating words deviating from normal practice, but like with all automatic procedures, the results were not perfect and needed some corrections during the linguistic annotation phase.

tor can only repeat the slightly apologetic rationale of Robinson and Solopova for including as much of textual detail as possible, even when it entails an increased chance of error:

We have no doubt that we have missed some tails and virgules which are there in the manuscripts, added some which are not in the manuscripts, and misplaced others. We could have eliminated all these categories of error simply by not transcribing tails and virgules. At a stroke, this would permit us to claim a considerably higher accuracy rate. However, we felt that it is better to give a transcription rich in detail, at the cost of some accuracy of rendering that detail, than a transcription which achieves perfection through impoverishment.

(Robinson and Solopova 2006: 15)

While the effect of accidental errors cannot by definition be ameliorated by pointing them out to the user, there are always some points in the transcription or annotation that the editor knows to be less than certain, and whose uncertainty should be made clear to the user in the name of honesty and scholarly integrity. For this reason, the present edition explicitly indicates those passages in the edition where the transcription of manuscript content is uncertain (see subsection 11.6.2) and analytical annotations which the editor is not fully confident in (see *Normalised forms* in subsection 11.9.1).

### 10.2.8 Identification of scribal hands

In the context of a linguistic edition, the documentation of the different scribal hands used in the manuscripts is essential for distinguishing textual elements originating from different time periods and different language users. While all of the recipe collections edited in the present edition have originally been written by a single scribe, all of them contain various kinds of emendations (see subsection 11.4.2) and annotations (see subsection 11.4.3) in different hands, and thus the separation of textual content written in different hands also separates the different stages in the production and use of the manuscript document. While the change of scribe always implies a change of hand, the converse is not true, as several hands—distinguished by differences in their formality, script type, or the physical medium (ink or pigment) used—can be associated with the same scribe. This means that the change of hand does not necessarily imply a change of scribe or even a separate stage in the writing process.<sup>48</sup>

As Symes (2004: 112) has observed in the context of medieval play texts, it can often be difficult to assign later changes and annotations made to a manuscript text to a certain hand when the changes or additions consist mainly of individual letters or words which provide very little to no information about the letterforms used by the annotator. This situation is also common in the *PD* manuscripts which contain a large number of marginal annotations consisting of isolated symbols or additions consisting of a single word or letter. In these cases the approach taken here is to

<sup>48</sup> While the addition of rubrication—which is by definition considered to be written in a different hand because of the different medium—was usually performed as a separate stage after the main text has been written, a scribe could well use a different, more formal hand for recipe titles even though he was writing them at the same time and with the same ink as the recipes themselves.

rely on what visual clues there are, namely the colour of ink and the width of the pen. This means that some of the separate hands defined in the descriptions of the manuscripts might in fact be variants of the same hand, written by the same scribe using different ink or a different pen. However, this has seen to be a lesser problem than that of conflating together several hands which might originate in different centuries and entirely different contexts.<sup>49</sup> The description of the different scribal hands as a part of the documentary annotation and their annotation in the transcription are described in subsection 11.4.1

## 10.3 'Non-editorial' annotation

As was argued in subsection 10.1.3 above, the layers of analytical annotation beyond the annotation of basic textual structure, expansion of abbreviations and possibly explanatory editorial notes are considered to be beyond the purview of the editor. While other layers of analytical annotation may naturally be provided by the editor—as in the present edition—these layers are not considered to be the product of the editorial process or the editor *as* an editor, but rather of separate scholarly endeavour by the editor acting as a linguistic, text-critical, historical, or literary scholar (see *Figure 10.1*). Unlike the editorial layers of annotation, which are considered to be useful, if not mandatory, for all types of digital documentary editions, the non-editorial annotation layers are motivated by the type of edited material and the intended use of the edition. In the case of the present edition, these motivators are its function as a *corpus-linguistic edition* and the nature of the edited texts as *multi-version discourse colonies*. As a response to these motivators, the present edition contains two such layers of annotation: the linguistic annotation establishing the lexical identity of the edited textual content, and the intertextual annotation of recipes found in the different *PD* versions that are considered to be *parallel versions* of each other.

### 10.3.1 Principles of linguistic annotation

The level of linguistically oriented annotation, on which the characterization of this edition as a *digital edition for corpus linguistics* largely rests, is based on the textual coordinate system described above, and consists of the explicit annotation of the *normalized form* and *basic word class* of each word-unit making up the edited manuscript texts. This solution makes it possible to simultaneously preserve the orthographic variation of the original text and to eliminate the difficulties it causes to corpus linguistic methods by, facilitating further linguistic analysis and annotation of the edited texts. Furthermore, as Johansson (2004: 103) points out, the association of all variant forms of a word with a shared referent, i.e. a normalized version, and thus with each other is in fact a prerequisite for the effective study of linguistic variation in manuscript texts.

Being analytical in nature, this kind of annotation is by definition tied to a certain theoretical view of language. In order to make the annotation as widely usable as possible, a conscious attempt has been made to keep the analysis involved in the annotation as elementary as possible, while still providing the re-

<sup>49</sup> The paucity of letterforms attested for a hand also makes it extremely difficult to date.

quired level of disambiguation. In practice this means restricting the linguistic categorisation of words to conventionally established and generic categories that are at least roughly acceptable within several different grammatical frameworks. Fortunately, the burden of universal applicability is somewhat lessened by the fact that the proposed normalized forms and basic word classes can be readily overridden either entirely or in part by defining a new analytic annotation overlay of the appropriate type.<sup>50</sup>

The fact that the linguistic annotation included in the present edition has been produced manually has both benefits and drawbacks. While the separate manual annotation of each individual occurrence of a word allows for a level of accuracy and functional distinction difficult to achieve with Middle English texts using algorithmic means, it is also—like any process of manual analysis of large amounts of data—“long and tedious”, as well as “open to human failings of inconsistency, misreading, and all kinds of error stemming from fatigue and boredom” (Francis 1980: 202). While the practical process of annotation was designed to minimize these problems and to maximize consistency, the manual nature of the annotation means that individual errors are bound to occur.<sup>51</sup> Despite the large amount of work involved in manual annotation, the result has been judged to be worth the effort, as manually annotated language corpora are not only useful for studying the annotated texts themselves, but also a vital prerequisite for developing learning linguistic algorithms, as they provide them with data from which to derive generalized rules from (Dipper 2008: 83-4). Since no automatic linguistic annotation algorithms have yet been developed for Middle English language varieties, linguistically annotated editions of such texts are potentially valuable also in this respect.

### Selecting normalized forms

In order to make corpus searches on the content of the edition more reliable and enable such standard corpus linguistic methods as keyword analysis and various kinds of type-token measurements, all of the variant spellings of what is considered as the same word need to be grouped together. This, of course, raises the issue of when are two linguistic tokens to be considered “the same word”, or a representative of the same *type*, i.e. on what level of granularity are these types defined? The two principal options available are the level of a *lexeme*, represented by a *lemma*, and the level of a *word form*, represented by the appropriate inflectional form of the word. As with many other issues, this edition adopts a policy of minimal intervention also with regard to the standardization of spelling variants, normalising them not into their *lemmas* but instead to standardized *morphological forms*, preserving the inflectional markers and other morphological characteristics

<sup>50</sup> While beyond the scope of this thesis, a TEI-based standard for such external annotation overlays is being developed by the author as a separate project based on this thesis.

<sup>51</sup> In order to annotate all instances of a single form (and related forms of the same word) consistently, all of the uniquely identified words contained in the six transcriptions were extracted from the edition along with their preceding and following context, and ordered alphabetically. The normalized forms of each word were then annotated to this concordance list, constantly referring to earlier forms of the same word in order to ensure consistency. After all occurrences had been annotated, the normalized forms and word classes were exported back into the edition based on the unique identifiers.

of the word but in a standardized orthography. Together with information on the basic word class of each word (see below), this provides for a level of word-sense disambiguation that is slightly above that of an untagged modern text, and should allow the lemmatisation, parsing and more detailed POS tagging of the texts using automatic taggers and parsing algorithms developed for modern texts.<sup>52</sup>

Of the two methods defined by Bowers (1989) for selecting the forms used to represent the shared *type* of variant forms, *regularization* and *normalization*, the present edition adopts the latter, defined as “imposing an external standard of regularity without the evidence of some specific precedent in the text being edited” (82).<sup>53</sup> As Bowers also points out, in the case of historical texts this in effect amounts to *modernization*. This approach, using the *Oxford English Dictionary* (OED Online) as the source for the normalized forms, was chosen in order to provide a reference point that is general and diachronically stable enough to be applicable also beyond the current edition and allows for diachronic comparison of texts from different periods.<sup>54</sup> This means that for those words of which a form is included in the *OED*, an appropriate inflectional form (or derivation, in some cases) of the closest *OED* headword is used. The capitalization of the normalized form reflects that of the original form, except for proper names, which have been capitalized regardless of the original form.<sup>55</sup>

For words which do not occur in the *OED*, the *Middle English Dictionary* (MED) has been used to provide the basis for the normalized form. The source of the normalized form is always indicated by the annotation, allowing an application to make automated queries to the relevant dictionary. Words that are too badly corrupted to be recognized and thus normalized are marked by replacing the normalized form by an element indicating a gap (see subsection 11.6.3). Nonce words or spontaneous formations that do not occur in the dictionaries but are readily understandable—being mostly combinations of an existing word and a prefix like “*ontrusse*” (meaning ‘to truss upon’), or names of dishes that are simply not recorded in the dictionaries (but occur more than once)—are given a ‘standard’ form based on the modern orthography of their components (“on-truss” in the case of the above example). The text has been linguistically annotated according to its *final* state represented by the manuscript. This means that words added later, either by the original scribe or subsequent annotators have also been annotated, and words whose lexical identity has been changed by word-internal emendations

<sup>52</sup> The differences between the syntaxes of Late Middle English and Present-Day English will naturally still pose challenges and make automated linguistic annotation less reliable than for modern texts.

<sup>53</sup> *Regularisation*, in contrast, is defined as “the bringing of inconsistent elements in a text into conformity by the adjustment of variants to some one regular form already present and assumed to be authorial” (Bowers 1989: 82).

<sup>54</sup> Since these modernized forms supplement the original forms instead of replacing them, the loss of data traditionally associated with normalisation does not occur and there is no reason to prefer regularisation to more ‘authentic’ forms over full normalisation.

<sup>55</sup> The various orthographic realizations of grammatical forms for which an equivalent still exists (and which do not have their own independent entry in the *OED*) have been normalized to their Present Day English form in order to facilitate automatic analysis of the text, while archaic forms for which no modern equivalent exists (and which thus have their own entry in the *OED*), have been normalized to the form given in the *OED*. This means that while the past participle form *baken* has been modernized to *baked*, the different realizations of the pronoun forms *thou*, *thee* and *thine* (as opposed to *you*) have *not* been modernized to *you* or *your*.

are annotated according to their emended lexical identity.<sup>56</sup>

### Determining word classes

Although corpus linguistic research has over the last decades problematized the assumption that “every word in every language can be consistently classified into a limited number of parts of speech” (Curzan and Palmer 2006: 31), the annotation of each lexical item in the edition with a word class was still considered to be a useful supplement for normalisation because of its ability in many cases to disambiguate between homomorphic forms, which are quite frequent in Middle English. While no word class taxonomy, no matter how rudimentary, can be considered to be ‘theory-neutral’, I share the belief of Lass (2004a) and Leech (2005) that it is possible to define “a ‘consensual’ set of categories on which people tend to agree” (Leech 2005) and which “nobody is going to quarrel with, or be unable to use” (Lass 2004a: 41), even if it does not completely agree with their preferred linguistic framework. In order to define such a categorisation based on the lowest common denominator between different frameworks, the eight word classes or parts-of-speech used in traditional school grammars, defined in terms of syntactic function and morphological patterns of grammatical forms (Atwell 2008: 503), were taken as a starting point. These categories—*noun*, *verb*, *adjective*, *preposition*, *pronoun*, *adverb*, *conjunction*, and *interjection*—are also used as the basic categorisation of words in many dictionaries, including the *OED*, and form the basis of more detailed tagsets like *CLAWS* and *NUPOS* (Mueller 2009), which means that they also serve as a preliminary step towards a more detailed categorisation according to one of these schemes.

In order to ‘update’ this categorisation and to align it with more recent linguistic practice, it has been extended with the category of *determiner*, used for example in the *Collins English Dictionary (CED)*—as well as the *CLAWS* and *NUPOS* tagsets—and covering words like *this*, *that*, *my*, *his*, *a*, *some*, *any*, etc. (Atwell 2008: 503).<sup>57</sup> In addition, to account for words whose class cannot be established, the classification here includes the meta-category of *unknown*. The categorisation used here thus corresponds to the one presented in Quirk et al. (1991: 67), with the exception that modal, primary and full verbs have not been distinguished but annotated simply as *verbs*, and the lesser category of *numerals* has been omitted, numerals having been annotated either as *determiners* or *nouns*, depending on their function (analogously to other quantifiers). The negative particle *not* has been annotated as an adverb, and the infinitival marker *to* has been categorized as a preposition even in this function, similarly to the *OED* and the *CED* (Atwell 2008: 508). In annotating multi-word lexical items which have been annotated as separate word-units according to the segmentation principles outlined above in subsection 10.1.3, the present edition follows the practice of the Pennsylvania University corpora and the Brown corpus in annotating them as sequences of several words (with the exceptions outlined in subsection 10.1.3; for example the item

<sup>56</sup> However, words marked for deletion in their entirety have still been annotated, as they can be considered to still be present, although latent, in the manuscript.

<sup>57</sup> The inclusion of the articles *a* and *the* in this category of determiners follows the practice adopted for the University of Pennsylvania linguistic corpora and differs from that of the Brown and LOB corpora which define a separate *article* category for them (Atwell 2008: 508).



as *well* would be annotated as a sequence of two separate adverbs.<sup>58</sup>

Just as for the selection of normalized forms, the primary practical reference used for assigning word classes to occurrences of words in the edition text has been the *Oxford English Dictionary* (*OED*) and the usage information contained therein. In cases where the base form of a word has been used, the word class defined for the item by the *OED* has been adopted. In the case of “homomorphs” (Quirk et al. 1991: 70), i.e. words that share the same morphological form but serve a different syntactic and semantic function and thus belong to different word classes, the appropriate word class has been selected based on the context and syntactic-semantic function of the word in the text and the usage examples given in the *OED*.<sup>59</sup> In the case of derivative forms of words (nominalisations, adjectival forms, etc.), the selection of a word class is based on the contextual function of the word, not on the word class of the parent word in the *OED*. The principal exception to the reliance on *OED*’s classification of words is the class of *determiners*, which is not used in the *OED* (where determiners are categorized mostly as adjectives). For this category, the definitions given in Quirk et al. (1991: 253–64) and Huddleston and Pullum (2002: 354–99) have been used as the basis for inclusion. Thus *articles*, as well as all of the different types of *quantifiers* (including cardinal numbers not used nominally) and *determinatives* defined in Huddleston and Pullum (2002: 358–99)—which correspond to the category of *determiners* (divided into *predeterminers*, *central determiners* and *postdeterminers* based on their position) in Quirk et al. (1991)—have been included in this category without subdistinctions.

It should be noted that similarly to all of the more elaborate part-of-speech tag-sets discussed by Atwell (2008), the present edition annotates each individual word on the basis of its contextual function, which means that the annotation of even adjacent word-units does not always occur on the same level, as words that are part of a constituent phrase are annotated according to their role within that phrase, while words operating as independent constituents on the sentence level are annotated on the sentence level. Thus the same lexical token can be annotated differently depending on the level of the syntactic structure it occurs on. As an example of the classification of such multiply homomorphic lexical items in the present edition, we can take the word ‘little’, which occurs in the edited texts as a *noun*, *adjective*, *adverb* and *determiner*:

- |        |   |             |
|--------|---|-------------|
| (1) a. | panne put a <i>litol</i> of þe white commode          | [noun]      |
| b.     | lete þem frye to gedre a <i>litol</i> while           | [adjective] |
| c.     | let hym kele a <i>litol</i> and lay hym in a vesselle | [adverb]    |

<sup>58</sup> This basic level of word-unit annotation will hopefully also facilitate the annotation of multi-word lexical items and other higher-level lexical structures by users focusing on the analysis of such structures, and lead to the accumulation of additional layers of analytical annotation.

<sup>59</sup> However, it has to be borne in mind that the process of *grammaticalization*, or the gradual transformation of content words into function words, means that the boundaries between word classes are not always clear-cut, and there are always words that lie somewhere between two word classes. The problem is especially acute in the case of historical texts, as grammaticalization is a diachronic process, and the modern classification of a word may not be appropriate for the Late Middle English period. This problem of diachronically shifting word classes also makes it impossible to determine word classifications that would be equally valid for all periods in the word’s history. For this reason, the word-classes assigned to ambiguous cases might occasionally differ from those that would be assigned to them in PDE.

d. powder of gingiuer and a *litol* vergus and salte [determiner]

In *Example 1a* the preposition *of* makes *little* the head of a noun phrase, while in *Example 1b* it qualifies the head, which in this case is *while*. In *Example 1c* *litol* serves as a ‘downtoning’ subjunct (Quirk et al. 1991: 597–602) for the verb *kele*, while in *Example 1d*, it serves as a quantifying determiner or a “degree determinative” (Huddleston and Pullum 2002: 393–5).

The classification of past participial forms is perhaps the most difficult and uncertain aspect of the word-class annotation. The basic guideline is that instances which retain their verbal characteristics and form a part of an attributive verbal phrase are annotated as verbs, and deverbalized forms that can be seen to behave as independent adjectives or heads of adjective phrases modifying nouns (either attributively or predicatively) are annotated as adjectives. Especially cases where the past participle of a mutative verb—i.e. an intransitive verb involving a change of place or state (Fischer 1992: 260)—occurs with the verb *be* (often in the subjunctive, e.g. *when they be boiled*) are problematic, since they can often be interpreted equally well as a perfect tense verbal form (‘when they have boiled’, with an auxiliary *be*), a predicative adjective (‘when they are boiled’, referring to the final state after boiling, with a copula *be*), or even a passive construction (‘when they are being boiled’, referring to the process. These kinds of cases have been judged on a case-by-case basis in the light of the basic principle mentioned above, preferring the adjectival interpretation unless contraindicated by the context (e.g. the presence of an adverb like *well* used only with verbs or a resultative adjective like *small*). While maximal consistency has naturally been the aim, it is recognized that subtle differences in the individual instances mean that it remains by necessity an ideal.<sup>60</sup>

### Treatment of foreign words

Although the issue of lexical borrowing and language mixing has been studied quite extensively, the question of distinguishing between different languages “raises major problems, both theoretical and practical, which remain largely unsolved” (Trotter 2000: 3), as was observed in chapter 6. Considering the dynamic state of linguistic development and essentially trilingual environment of these texts, it is obviously extremely difficult—if not impossible—to distinguish between ‘Middle English’ words and ‘foreign’ words. Late Middle English was already in itself a hybrid of Old English, Anglo-Norman and Latin, and in the 15<sup>th</sup> century it was still constantly adopting new words from Latin and French, which makes drawing the line between foreign words and loans already appropriated into the English lexicon very difficult. Although the etymological classification of culinary lexis is beyond the scope of this work, the regularization of spelling variants described above under *Selecting normalized forms* requires that even words that do not belong to the Middle English lexicon—as represented by the *OED* and the *MED*—are assigned a standard form.

<sup>60</sup> Word order has not been considered a decisive factor, as for example Middle English recipe titles—being heavily influenced by French—very frequently use adjectives also in postmodifying positions. For this reason, the adjectival interpretation has been chosen for recipe names and other nominal constructions, regardless of the word order (*venison baked* vs. *baked venison*).

In order to reflect the hybrid and expanding nature of Late Middle English, this edition takes a rather inclusive view of Middle English and considers all words that *both* have a base form that appears in either the *OED* or the *MED* as an independent lexical item with the appropriate meaning and a pre-modern dating, *and* conform to Middle English syntax and morphology to be an established Middle English word.<sup>61</sup> This means that only those words that do not fulfil these requirements are annotated as foreign, using the method described in *Foreign-language words* in subsection 11.9.1. In practice, this includes:

- 1) words that do not have a base form occurring in either of these dictionaries with an appropriate sense assigned to them before c. 1800,
- 2) words inflected according to the paradigms of a language other than English, including e.g. adjectives inflected according to number or nouns with Latin case markers,
- 3) words determined by a foreign determiner (e.g. “*la*”, “*un*”, “*autre*”) *and* are found in a dictionary of that language, or
- 4) words represented in the *OED* only by modern examples used in the context of borrowed foreign expressions (e.g. “*de*” or “*cetera*”).

Although this kind of word-level annotation of language can be argued to be inconsistent and ‘messy’, since it frequently results in phrasal units like names of dishes or ingredients being analyzed as partially Middle English and partially foreign, it has been adopted precisely for this reason: it helps to highlight the hybrid and indeterminate linguistic status of many of these phrases and of Middle English itself. It should also be noted that the annotation of these words as ‘foreign’ does not necessarily mean that they would have been considered especially foreign by contemporaries, but rather that they did not subsequently establish themselves in the English lexicon.<sup>62</sup> As such, this annotation should not be seen as an entirely reliable analysis of code-switching but rather as a practical measure necessary for establishing a normalized lexical identity for these words. The normalized forms of foreign words are established similarly to those in English, using standard dictionaries as the reference point. For French, the source used for the normalized forms is the *Dictionnaire de l’Académie française* (DAF). For Anglo-Norman words which do not occur in the DAF, the *Anglo-Norman Dictionary* is used, similarly to the *MED* for English (with the exception that Anglo-Norman words are annotated separately from Middle French ones). For Latin, the authority corresponding to the *OED* is the *Oxford Latin Dictionary* (OLD), while the Latin dictionary of Lewis and Short (1879) has been used similarly to the *MED* for medieval words that do not occur in the *Oxford Latin Dictionary*.

The foreign languages represented in the texts of this edition are Latin and (Anglo-Norman or Old) French.<sup>63</sup> In ambiguous cases the distinction between these two languages is made using similar criteria as those used for Middle En-

<sup>61</sup> However, even the *MED* is not always consistent in its inclusion or exclusion of words based on their linguistic status, some words which could arguably be seen as Latin being included as ME *hapax legomena* while others are excluded (Schendl 2000: 86-7).

<sup>62</sup> This means that the annotation used here is not directly comparable to the practice suggested by for example Burnley (2001: 26), which counts as foreign only those words which are likely to have been considered foreign by the original audience of the work.

<sup>63</sup> In addition to which also modern English occurring in various modern annotations in the MSS is annotated to separate it from Middle English.

glish words, apart from the fact that here contextual consistency is used to decide between ambiguous cases.<sup>64</sup> In practice this means that those foreign words that have a base form found in the online version of either the *DAF* or the *AND* and can be judged to conform to Old French or Anglo-Norman morphology are considered French, while those that have a base form found in the *OLD* or in Lewis and Short, and can be judged to conform to (medieval) Latin morphology are considered Latin.<sup>65</sup> Words whose language cannot be identified are annotated as “undetermined” as per the International Organization for Standardization (ISO) 639-3 standard (SIL International 2011). The distinction between ‘French’ and ‘Anglo-Norman’, which is in most cases impossible to make since most words can be shown to occur in both the continental and Anglo-Norman varieties, has here been made in a way that prioritizes French, annotating as Anglo-Norman only those words which are not attested in the *Dictionnaire de l’Académie française* and can thus be considered to be exclusive to the Anglo-Norman variety, all of the words shared by these varieties being annotated simply as French.

### 10.3.2 Intertextual annotation of parallel recipes

While the linguistic annotation described above is associated with the identity of the present edition as a specific *kind* of edition, the annotation of intertextual relationships between recipes that are considered to be ‘the same’ or *parallel versions* of each other is related to the nature of the edited text as a group of related discourse colonies. The central concept in the relationships between recipes is that of *parallel recipe* or *parallel version*, which is here used to refer to recipes—occurring either within a single manuscript or, more commonly, in different manuscripts—that describe essentially the same procedures using more or less the same ingredients (allowing for the occasional omission or addition of an individual ingredient or minor procedure), although not necessarily expressed in exactly the same terms or syntactic structures. While the criteria for recognising parallel recipes are necessarily subjective and require the application of critical judgement, in practice there is rarely any doubt as to whether two recipes are parallels of each other, at least in the case of the current manuscripts— if there is considerable uncertainty, the recipes are not parallels of each other.

In order to allow the study of the structural and textual variation within the family, these parallel versions are linked together by an annotation overlay consisting of links associating all of the parallel versions for each of the 371 unique recipes found in the *PD* family to each other. Vanhoutte (2004: 168) has described a very similar method of accounting for textual variation in multi-version texts, calling it the “linkemic” approach, based on the interlinking of parallel variant versions in different versions of the text. The basic unit in this system is a *linkeme*, which is defined as the smallest unit, either structural or semantic, that is used for linking together the different variant versions of the text. Whereas the genetic edition of *De teleurgang van den Waterhoek* by Stijn Streuvels prepared by

<sup>64</sup> If a word—such as the preposition “a”—could equally well be judged to be French or Latin, the decision is made based on the language of the surrounding text: French in the case of with a Middle English or French context and Latin only within an otherwise Latin context.

<sup>65</sup> For assigning the normalized form to foreign words, the same procedure described above for Middle English words is used, using the *DAF*, *AND* and the *OLD* in place of the *OED* and *MED*.

De Smedt and Vanhoutte (2000) using the linkemic approach uses paragraph as the linkeme, the present edition considers individual recipes to constitute suitable linkemes on the level of the entire recipe collection.<sup>66</sup>

This annotation is based on the close reading of the recipes in each of the six versions, as well as the work of earlier scholars,<sup>67</sup> and it is used in chapter 13 for analysing the internal relationships and textual history of the six members of the *Potage Dyvers*. The annotation and analysis of the relationships between the recipes of the six *PD* versions, together with the preliminary overview of their relationships to other families of recipe collections presented in section 9.3, is intended to form the first step in the mapping of the English culinary recipe tradition in terms of the transmission and transformation of recipes from collection to collection and from century to century. This kind of an annotation between different versions of a discourse colony means that the present edition is not just “an edition of all the texts in all of the [...] witnesses” (Robinson 1993: 10), but also “an edition of all the texts which those witnesses collectively produce: not just an archive, as it would be if each transcription were isolated from all of the others, but an edition which gives us analytical leverage that we could not otherwise attain” (Flanders 2009: 61).

This intertextual annotation is also intended to form the basis for a large diachronic corpus of English recipes from the Middle Ages to the present day, linking together parallel recipes not only within the *Potage Dyvers* family but across the whole corpus. The future annotation of not only relationships between the parallel recipes, but of similar relationships on a practical rather than textual level between different recipes describing a version of the same basic dish, across this corpus will allow us to trace the transmission and adaptation of recipes throughout from collection to collection throughout the centuries. This kind of an intertextually annotated collection of digitally edited recipe collections would constitute a “dense research matrix in which individual items connect with one another in sophisticated ways” (Gants 2006: 123), linking each individual edition into “one vast textual field” (Flanders 2006: 148):

[I]ndividual editions will benefit if they can be treated not only as distinct units but also parts of larger collections. An edition of *Macbeth* should interact with other plays by Shakespeare, with all Elizabethan and Jacobean drama, and with all online dramatic texts in any language. The more powerfully individual editions and digital libraries interact with documents that accumulate over the coming years and decades, the more useful they will be as a whole. (Crane 2006: 289)

## 10.4 Conclusion

The fundamental editorial principle of the present edition could be summed up as an attempt to combine the detailed description of the textual content and struc-

<sup>66</sup> In a similar fashion, the individually annotated word-units could serve as linkemes within the individual recipes, once they are linked together through the collation of all the variants of each unique recipe.

<sup>67</sup> Mainly (Hieatt 2004) who identified the relationship of MS Ad to the edition of Austin (1888) (based on MSS As, D, H279 and H4016).

ture of the edited documents with an equally detailed description of their documentary (i.e. visual, palaeographical and physical) features in a way that relates these two levels of the textual object to each other and allows the versatile analysis and continued annotation of the edited text. In order to accomplish this, a faithful graphemic transcription of the textual content is combined with descriptive annotation encoding the physical structure and condition, as well as various visual features of the document, to produce a *model* of the original manuscript as a physical document. This descriptive annotation is supplemented by a detailed documentation of the bibliographical, codicological and historical features of the original document, and a layer of analytical annotation representing its textual structure. This analytical annotation describes the text as a hierarchical structure extending from the level of the entire recipe collection to the level of individual word units, and also serves as a textual coordinate system to which further analytical annotation, stored in separate *annotation overlays*, can be linked. The addition of further layers of analytical annotation is demonstrated by the addition of a linguistic annotation overlay providing each word in the edited text with its normalized form and basic word class, and an intertextual overlay linking together all parallel versions of the same recipe across all six MSS.

The text of the six versions of the *PD*, including all associated material such as tables of contents and bills of fare, is transcribed graphemically as it appears in the manuscript, including all abbreviation markers and other special symbols. The combination of diplomatic transcription principles and the annotation of the physical structure of the manuscript to the annotation of the logical textual structure means that the transcriptional approach of the present edition can be characterized as a hybrid between a *diplomatic transcription* and a *linear transcription*, using the terminology defined by Pierazzo (2009: 172) for describing the transcriptional principles of genetic editions. It resembles the former in that “the text is transcribed with the intention of reproducing in the transcription all the features of the source document, including existing punctuation, marginal insertions reproduced in the exact position where they occur in the original, special characters and so on”, reproducing the exact appearance of the *document*, while simultaneously reconstituting the logical structure of the *text*, like the latter. In essence this hybrid or compound approach to transcription means that the text is represented on three different levels: graphemic, expanded, and normalized.<sup>68</sup> This multi-layered approach to transcription and annotation, while entailing some compromises in terms of simplicity and elegance of the annotation and its processing, produces an edition which can not only be combined with different analysis and presentation technologies to produce a variety of different editorial outputs to serve a variety of purposes, but also continue to increase in usefulness as new layers of analytical annotation are added to it.

<sup>68</sup> In this regard these transcription guidelines of the present edition resemble those of *The manuscript texts from Vadstena monastery in digitized form* project, described by Johansson (2004: 98), which also transcribes the text of the manuscripts on two levels which they call the “facsimile” and “diplomatic” levels (corresponding to the graphemic and expanded levels of the present edition) and also explicitly annotates each word-unit in the text. The main differences in their approach are the incorporation of editorial emendations to the diplomatic/expanded transcription and the analytical annotation of each word-unit with the appropriate *lemma* instead of a normalized morphological form. Unlike the present edition, the project also uses the annotation developed by the *MENOTA* project instead of native TEI P5 markup.

In addition to the production of the present edition, the editorial principles described here are intended to serve as the basis of a more general set of guidelines for editing historical documents, and have thus been designed for the annotation of not only those things that are known to be useful in the context of the present edition and the future researched questions envisioned for it, but also a variety of things that are *potentially* significant and useful for the wider intended uses and users of such editions. Additionally, the present edition is also intended to serve as the starting point for a long-term project of compiling a diachronic corpus of English recipe texts one edition at a time. This means that the editorial principles have been designed to allow not only the transcription and annotation of texts from different historical periods, but also to facilitate and encourage such piecemeal production of textual corpora by individual scholars or small projects, as advocated earlier by Honkapohja, Kaislaniemi and Marttila (2009).





# Chapter 11

## Annotation practices

The process of documenting markup choices is actually one of learning about a body of texts. Specialists know the text but not from the standpoint of imposing markup from a fixed set of elements such as the TEI guidelines. And it is not always obvious which elements should be used for encoding, even if there is agreement about what to encode. (Durusau 2006: 303)

As the above quotation from Durusau points out, editing in the digital medium requires the editor to have an understanding not only of scholarly editing, but also of the use of *text encoding* for the representation of various textual and paratextual features of the original document (Bøe, Jørgensen and Taugbøl 2004: 60). It also means that any decisions made in encoding and annotating a text are far from obvious, self-evident or ‘natural’, and that it is therefore unreasonable to expect them to be automatically understandable to a user or repeatable by another annotator without an explicit documentation of the rationale behind them. In discussing the concepts of *interchange* and *interoperability*, Bauman (2011) has also emphasised the importance of the “contemporaneously generated detailed prose documentation” of the annotation applied to the text, arguing that it “can significantly facilitate the use of the text in future systems and processes, both those we can predict and those we cannot, both by those that generated the text and by unforeseen recipients of it”. Since fostering future research and allowing its results to be shared with others is a central concern for the present edition, the task of making the annotation as understandable and transparent as possible for its potential users is also considered to be of first priority.

While chapter 4 presented the general theoretical framework for the present edition and the previous chapter outlined the editorial principles underlying the transcription, encoding and annotation of the edited documents, this chapter will provide a detailed account of the specific ways in which these editorial principles have been implemented in the annotation of the base data files and annotation overlays which make up the present edition as a *data archive*. While these annotation practices—like the editorial principles described in the previous chapter—are

primarily descriptive of the present edition, they are also intended to be programmatic in the sense of laying the basis for the development of a practical annotation scheme that would allow also other scholars and editorial projects to create digital editions of historical documents that would be useful for the purposes of corpus linguistic study. With this in mind, they have been designed to be easily adaptable for other types of historical documents with only minor modifications and expansions.<sup>1</sup>

As was explained in section 5.7, the annotation used in the present edition is implemented using eXtensible Markup Language (XML) and is based on the *Text Encoding Initiative's Guidelines for Electronic Text Encoding and Interchange* (TEI Consortium 2014). However, as Lehmborg and Wörner (2008), among others, have observed, practical experience of the general *TEI Guidelines* has showed that “more streamlined tag sets and tighter restrictions with respect to tag content” are required “in order to enhance the usefulness of automatic processing and eliminate ambiguity” (499). While the annotation described here makes some minor extensions to the *TEI Guidelines* and is therefore technically a *TEI Extension*, it has nevertheless been designed to be as close to TEI Conformant as possible, and significant effort has been put into implementing the features required by a corpus-linguistic digital edition within the specifications of the *TEI Guidelines*. This means that apart from some individual details, it is essentially a restricted subset of the *TEI Guidelines* that excludes a large number of the elements and attributes defined by the guidelines and severely restricts the values of many attributes to facilitate the consistent annotation of manuscript features.

The subset of *TEI Guidelines* that makes up the basis for the annotation described in this chapter, as well as the further restrictions and extensions made to them, is formally documented by a TEI ODD (TEI Consortium 2014: 643-67) document and the corresponding Relax NG Schema (Clark and Muraka 2001) included in appendix A. However, since both the XML metalanguage and any formal schema used to govern its use are concerned exclusively with the *syntax* of the annotation, the *semantic* aspects of annotation, i.e. its mapping to the original document, needs to be documented separately (Buzzetti 2009: 53). This chapter is therefore intended to provide the semantic component of the TEI customisation used for encoding the present edition, standing in the same relation to the written *TEI Guidelines* as the formal Schema stands to the formal TEI XML specification, being simultaneously an extension and a selective reduction of it.

Since the practical annotation process does not usually follow the hierarchy of annotation layers described in section 10.1, the description of the different types of practical annotation structures used to realise it are introduced in a slightly different order, more relevant for the process of annotation and to the structure of the edition as a *data archive*. In terms of the physical organisation of the data on the level of file structure, the present edition consists of a single TEI XML *base data file* for each of the six *PD* manuscript versions, and a selection of associated *annotation overlays*, similarly encoded as TEI XML files.<sup>2</sup> The majority of this chapter is dedicated to the description of the internal structure of the *base data files*, which

<sup>1</sup> This work of adapting and generalizing the solutions developed for this edition for a wider variety of text types is planned in the context of the Digital Editions for Corpus Linguistics (DECL) project (for more information, see <<http://www.helsinki.fi/varieng/domains/DECL.html>>).

<sup>2</sup> The names of these files and their location on the included CD-ROM is described in appendix A.

contain the *data* of the edition, i.e. the transcription and the descriptive annotation of the original document, while the structure of the included annotation overlays is described in sections 11.7, 11.8 and 11.9.

The first section of this chapter describes the structure and content of the *TEI header* containing the documentary metadata describing the edited document and its relationship to the *analytical model* constituted by the data archive. The data archive itself is first described in terms of the annotation of textual structure which forms the primary organizational principle of the textual model as an Ordered Hierarchy of Content Objects (OHCO). This is followed by an account of the descriptive annotation layers, beginning with the physical structure of the original document and its relation to the textual structure, and the annotation of the process of ‘textual genesis’ (scribal hands, emendations and later annotation). The next two sections describe the annotation of the visual characteristics of the document (highlighting and decoration, graphical elements, and special symbols) and of the physical condition of the original document (damage and illegibility). The last three sections move towards analytical annotation, the first describing the annotation of manuscript abbreviation and its editorial expansion, while the last two describe the annotation overlays containing editorial explanatory notes (section 11.8) and the ‘non-editorial’ or research-based annotation of lexical forms and intertextual relationships (section 11.9).

## 11.1 Metadata header

Each transcription in this digital edition is accompanied by a large amount of *documentary annotation*, consisting of metadata describing not only the transcribed manuscript, but also the electronic text itself, the ways in which data has been encoded into it, and its revision history. This metadata is contained in what the *TEI Guidelines* call the *TEI header*, “an electronic analogue to the title page attached to a printed work” (TEI Consortium 2014: 18). In the present edition, the *TEI header* consists of four principal components, defined in the *TEI Guidelines*:

- 1) a *file description*, represented by a <fileDesc> element, containing “a full bibliographical description of the computer file itself” (TEI Consortium 2014: 18), and perhaps even more importantly, “information about the source or sources from which the electronic document was derived” (18), i.e. the original manuscript. The structure and contents of the file description are explained below under subsection 11.1.2, and those of the *manuscript description* contained within it separately under subsection 11.1.6.
- 2) an *encoding description*, represented by an <encodingDesc> element, which is used to describe “the relationship between an electronic text and its source or sources” (TEI Consortium 2014: 18). In this edition, the encoding description contains a concise summary of the information contained in this and the previous chapter. The structure and contents of the encoding description are explained below under subsection 11.1.3.
- 3) a *text profile*, represented by a <profileDesc> element, containing classificatory and contextual information about the transcribed text. In the context of this edition, the text profile is used to provide information about: a) the languages used in the manuscript text, with approximate proportions based on

the text-structural annotation of language (described in subsection 11.2.7 below), b) the subject classification of the edited text using Library of Congress keywords and class codes, c) the corpus-linguistic ‘situational parameters’ of the edited text, which can be used as the basis of categorising the edited text within a linguistic corpus.

- 4) a *revision history*, represented by a <revisionDesc> element containing a series of records that describe the changes made to the document over its production history. This part of the header is intended mainly for the purposes of version control and documentation of the production process of the edition. The format of the revision history is explained below under subsection 11.1.5.

Of these components, most of the *file description*, the entire *text profile* and to an extent also the *revision history* are particular to each text in this edition, while the *encoding description* is by definition identical for each text that follows the same guidelines. The file description also contains several components that pertain to the entire edition (distribution, licensing, etc.) and are therefore identical for all six texts. The revision histories of all the texts are also very similar, although not identical.

### 11.1.1 General annotation of metadata

In addition to the above mentioned structural components and their subcomponents, the *TEI header* makes use of a selection of general phrase-level metadata elements defined by the *TEI Guidelines* for the annotation of names (TEI Consortium 2014: 85-8, 305-6, 419-30), dates (93-4, 418-9), measurements (90-3), manuscript *loci* (303-5), quoted and mentioned text (76-7), technical terms (76-7), hyperlinks (96-9), and for the documentation of TEI XML elements themselves (644-5). This phrase-level annotation of individual pieces of data is intended to explicate the information—both quantitative and qualitative—contained in the header and to facilitate the automatic mining of this data for the purposes of data aggregation and visualization, and to allow the metadata to be used to filter and annotate the results of corpus searches. Since these elements are used within all of the component parts of the TEI header, they will be described first.

#### Names

For the documentation of the people involved in the production of both the digital edition and the original manuscript, as well as references to geographical places, the following elements defined in the module for *Names and Dates* in the *TEI Guidelines* (TEI Consortium 2014: 416-58) are used:

- <persName>** (personal name) contains a proper noun phrase referring to a person, with its component parts encoded using the <forename>, <surname>, <addName>, <nameLink>, <placeName>, <genName> and <roleName> elements and the @key attribute providing a standardised form of the name;<sup>3</sup>

<sup>3</sup> For the <persName> elements representing persons involved in the production of the edition, i.e. the editor, an @xml:id attribute is also used to provide a means of referring to the name and indicating

<b>&lt;orgName&gt;</b>	(organizational name) contains the name of an organization;
<b>&lt;forename&gt;</b>	contains a given or baptismal name of a person;
<b>&lt;surname&gt;</b>	contains the family name of a person;
<b>&lt;addName&gt;</b>	(additional name) contains an additional name component, such as a nickname or alias;
<b>&lt;nameLink&gt;</b>	(name link) contains a connecting particle “used within a name but not regarded as part of it” (TEI Consortium 2014: 420), such as <i>of</i> or <i>de</i> ;
<b>&lt;placeName&gt;</b>	(place name) contains a geographical place name, used either as a part of a personal name or as a geographical reference;
<b>&lt;genName&gt;</b>	(generational name component) contains a name component used to distinguish between similar names on the basis of age or generation, such as <i>Jr.</i> or <i>III</i> ;
<b>&lt;roleName&gt;</b>	contains a name component that indicates a particular social role, such as an official title or rank.

## Dates

For the encoding dates and date ranges in a standard form that can be automatically processed, the `<date>` element is used with the attributes `@when`, `@from` and `@to` (TEI Consortium 2014: 93-4). The prose representation of the date or date range, in whatever form required by the textual context, is contained within the `<date>` element, while a normalized form of it is represented as the value of the `@when` attribute (for dates) or the `@from` and `<to>` attributes (for date ranges). For approximate dates (such as the dating of hands or other features by century), the attributes `@notBefore` and `notAfter` are used.<sup>4</sup> The values used for the attributes are instances of the XML Schema datatypes `'date'`, `'gYearMonth'` and `'gYear'` (Malhotra and Biron 2004). (E.g. `<date notBefore="1500" notAfter="1599">the 16th century</date>` or `<date when="1677-08-03">3 August 1677</date>`.)

---

responsibility for its annotation within the annotated text.

<sup>4</sup> In expressing approximate date ranges like centuries or their subdivisions, the century is considered to begin on ‘year 0’, i.e. ‘the 15<sup>th</sup> century’ is considered to cover the years 1400–1499. Correspondingly, also quarter and half centuries are considered to begin on an even year and end on an odd one in order to avoid adjacent ranges overlapping on the even dividing year. Since many of the datings are very uncertain, an inclusive approach to date ranges is used, the terms *early*, *mid* and *late* century being understood loosely as referring to half-centuries and covering years 0–49, 25–74, and 50–99 of the century and thus overlapping (i.e. a manuscript dated as mid-15<sup>th</sup> c. is considered more likely to be written after an early 15<sup>th</sup>-c. one, but the possibility is left open that they were in fact written around the same time, i.e. s. xv<sup>2</sup>). Quarter centuries, e.g. s. xv<sup>3</sup> in traditional notation, are understood similarly, the example being understood as referring to the years 1450–1474. Similarly, the turn of the century (e.g. s. xiv/xv) is also understood loosely as a 50-year period from 1375 to 1424. In terms of searching this means that a search for manuscripts from 1400–1449 (i.e. ‘the early 15<sup>th</sup>-century’) would return all manuscripts whose possible dates overlap with this range, including for example ones dated as s. xiv/xv, *early 15<sup>th</sup> c.*, *mid 15<sup>th</sup> c.*, s. xv<sup>1</sup>, 1417, 1405–15 or 1440–1470.

## Measurements

Various kinds of physical measurements, used mostly in the description of the manuscript as a physical object, are encoded using the generic elements `<measure>` (for amounts of e.g. folia) and `<dim>` (for dimensions), and the more specific elements `<height>`, `<width>` and `<depth>` (TEI Consortium 2014: 90-3, 301-2). Similarly to dates, the contents of the element are made up of a free-form prose description of the measurement, and its normalized form is encoded using the attributes `@unit` and `@quantity`.<sup>5</sup> In the case of measurements containing variation, the attributes `@min` and `@max` are used in place of `@quantity` to encode the minimum and maximum values of the range. (E.g. `<height unit="mm" min="145" max="155">145-155 mm</height>`.)

## Manuscript *loci*

In cases where certain features or phenomena pertain to a specific part of the manuscript, this *locus* is encoded in a normalized form using the `<locus>` element defined by the *TEI Guidelines* (303). As with dates and measurements, the element is placed around the prose representation of the manuscript locus, most commonly a range of folia, and the `@from` and `@to` attributes are used to indicate the range using the `@n` attribute values of the appropriate `<pb>` elements in the edition (see subsection 11.3.1).<sup>6</sup> If the reference of the element is to a single manuscript page, the `@to` attribute is omitted.

## Quoted and mentioned text

Text in the TEI header that has been quoted either from the edited text—or from other sources (indicated by a pointer to a bibliographic reference)—has been enclosed within the `<quote>` element (TEI Consortium 2014: 71). If no `<ptr>` element referring to a bibliographic source is provided, the quotation is from the edited manuscript text itself. Characters, words or phrases that have been “mentioned, not used” (TEI Consortium 2014: 76) are enclosed within a `<mentioned>` element. No attributes are used for either of these elements.

## Technical terms and foreign words

Special technical terms—such as would conventionally occur in italics in a printed text—are indicated by enclosing them within the `<term>` element (TEI Consortium 2014: 76). The formula used for indicating the collation of the manuscript (e.g. *1*<sup>6</sup>, indicating that the first quire of the MS consists of six leaves) has been annotated using the element `<formula>`, and heraldic blazons describing coats of arms in the MS description have been annotated using the element `<heraldry>`. Words or phrases that are not English and would conventionally be printed in italics for this reason, are annotated using the `<foreign>` element with an `@xml:lang` attribute

<sup>5</sup> The `@type` attribute is used in some cases to characterize the dimension, e.g. “line-height” or “margin-bottom”.

<sup>6</sup> For parts of the manuscript not included in the edition, the folio numbers are extrapolated based on the foliation used for the edition.

indicating the language using the three-letter code defined for the language in (SIL International 2011).

### Hyperlinks

Hyperlinks, whether pointing to elements within the document itself or online, are represented using the <ptr> and <ref> elements (TEI Consortium 2014: 96-7).<sup>7</sup> These elements take two or three attributes, depending on their type. The first of these is the @type attribute, which takes one of the following four values: 'biblref' for bibliographical references to works cited in the source bibliography, 'handref' for references to a scribal hand described in the *manuscript description*, 'online' for hyperlinks to an online resource on the Internet, and 'crossref' for internal cross-references to some other section of the TEI header. The second attribute shared by all <ptr> and <ref> elements in the header is the @target attribute, which contains a Uniform Resource Identifier (URI) pointing to the relevant object. The most suitable treatment of this URI by processing applications depends on the type of hyperlink and the intended use.<sup>8</sup> In addition to these attributes, bibliographic references also make use of the global @n attribute to indicate a page or page range of the work to which reference is made.

### Text formatting

Special formatting—primarily italicization and superscripting—either found in the sources quoted in the TEI header or dictated by convention (such as the presentation of the number of folia in a collation as superscript) is indicated in the header by using the @rendition attribute with a value defined in the <tagsDecl> component of the <encodingDesc> (see subsection 11.1.3) on the appropriate element.<sup>9</sup>

### TEI XML elements

In documenting the annotation used in the edition, the TEI header frequently makes reference to the various XML elements defined in the *TEI Guidelines*. For this purpose, the header makes use of the <gi> (generic identifier) element, containing the name of the element referred to. Similarly, the names of element attributes referred to are enclosed within the <att> (attribute) element and any attribute values within a <val> (value) element (TEI Consortium 2014: 644-5).

#### 11.1.2 File description

In the context of this edition, the principal part of the TEI header is the *file description* (TEI Consortium 2014: 23–35) which is used to encode the ‘traditional’ bibliographic information about both the digital edition and its original source, i.e.

<sup>7</sup> The only difference between these elements is that <ptr> is an empty element, while <ref> can contain textual content describing the link.

<sup>8</sup> For example, in a HTML rendering, a bibliographical reference might be replaced with a hyperlink to a bibliography file, its contents made up of a traditional citation like (e.g. author-year), while an online hyperlink would be replaced by a hyperlink with the Uniform Resource Locator (URL) address as the content within angle brackets.

<sup>9</sup> If no suitable element is present, the generic <hi> element described in subsection 11.5.1 is used.

the manuscript. The `<fileDesc>` element—whose hierarchical structure is outlined in *XML Example 1*—consists of the following four major and two minor subcomponents, defined by the *TEI Guidelines*:

- 1) `<titleStmt>`—the *title statement*
- 2) `<editionStmt>`—the *edition statement* (minor)
- 3) `<extent>`—the *extent* of the file (minor)
- 4) `<publicationStmt>`—the *publication statement*
- 5) `<notesStmt>`—the *notes statement*
- 6) `<sourceDesc>`—the *source description*

---

**XML Example 1:** *The XML document structure of the file description in the TEI header.*

---

```
<fileDesc>
  <titleStmt>
    <title><!-- The title of this digital edition. --></title>
    <editor>
      <!-- The name of the editor responsible for the production of this edition. -->
    </editor>
    <sponsor>
      <!-- The name of the organization under whose auspices this edition has been produced. -->
    </sponsor>
    <funder>
      <!-- The name of an organization or institution that has funded the creation of this edition. -->
    </funder>
  </titleStmt>
  <editionStmt>
    <edition n="1">
      <!-- Information about the current release of the edition. -->
    </edition>
  </editionStmt>
  <extent>
    <!-- The word count of the edition. -->
  </extent>
  <publicationStmt>
    <authority>
      <!-- The authority responsible for the release of the edition. -->
    </authority>
    <date>
      <!-- The release date of the edition. -->
    </date>
    <distributor>
      <!-- The organization through which the edition can be obtained. -->
    </distributor>
    <idno type="filename"><!-- The file name of the document --></idno>
    <availability>
      <!-- Information about the availability (repository and license) of the edition. -->
    </availability>
  </publicationStmt>
  <notesStmt>
    <!-- A series of <note> elements referring to specific points in the edition. -->
  </notesStmt>
  <sourceDesc>
    <msDesc>
      <!-- The manuscript description (see "Manuscript Description" below) -->
    </msDesc>
  </sourceDesc>
</fileDesc>
```

---

The *title statement* “groups information about the title of the work and those responsible for its intellectual content” (TEI Consortium 2014: 24). Within it, a `<title>` element is used to contain a title identifying the *electronic edition* of the text (not the original manuscript text), while the `<editor>` element contains the



name of the editor of the text (annotated with a `<persName>` element<sup>10</sup>), who in this context is considered to be responsible for all aspects of the digital edition (transcription, proofreading, annotation and metadata). The institutional context of the work is documented by a series of `<sponsor>` and `<funder>` elements, the former identifying the organizations or institutions (e.g. the university departments and research units) under whose auspices the edition was created, and the latter the individuals, institutions or organizations that provided funding for the editorial work.

The *edition statement* is a minor component of the *file description* containing a single `<edition>` element that briefly characterises the particular *edition* (in the sense of ‘version’ or ‘release’) of the digital text represented by the file. A numeric identifier for the edition (“1” for the edition released with this thesis) is provided by the `@n` attribute of the element. The `<extent>` element is another minor but important part of the file description, and indicates the size of the electronic text in the form of a total count of word-level units (including words and numbers, see subsection 11.2.5) contained by the main `<text>` element of the edition for the purposes of verifying the integrity of the file.

The *publication statement* contains information about the publication and distribution of the digital edition. Since the digital edition is not officially published in the conventional sense, the `<authority>` element, representing the “agency responsible for making an electronic file available” (TEI Consortium 2014: 25) is used instead of the more conventional `<publisher>` and `<pubPlace>` elements to indicate the *release authority* through which the text is made available, along with a `<date>` element indicating the release date of the edition. Regarding the availability of the edition, the `<distributor>` element is used to identify the organization through which the edition will be available, in the case of the present edition, the University of Helsinki. The `<availability>` element contains information on where to obtain a copy of the digital edition, along with the license under which it is released, organized as paragraphs of annotated prose text.

The *notes statement*, which is defined in the *TEI Guidelines* for gathering together “any notes providing information about a text additional to that recorded in other parts of the bibliographic description” (TEI Consortium 2014: 31), has been slightly re-purposed in this edition to be used as the container for a series of textual notes that provide information about the textual and documentary characteristics of individual *loci* in the text. Each of these notes is represented by a `<note>` element with the following attributes:

- @type** identifies the note as a *textual note* ('textnote');
- @xml:id** provides a unique identifier for the note, which can be used to refer to it;
- @target** points to the `@xml:id` values of those elements within the main `<text>` of this document that this note applies to;
- @resp** points to a `<persName>` element defined in the header, representing the person responsible for the note (here invariably the editor).

<sup>10</sup> Unlike other `<persName>` elements, which simply use the `@key` attribute to assign a canonical form for the name to establish its identity, this `<persName>` element (as well as any others representing a person involved in the production of the edition) uses the `<xml:id>` element to provide it with a unique identifier that can be referred to by the `@resp` attribute of other elements in the edition (see subsections 11.6.2, 11.6.3, *Normalised forms* in subsection 11.9.1 and section 11.8).

The *source description* is the largest single component of the *file description*, containing a detailed description of the source document of the digital edition, i.e. the original manuscript. This description is presented in the form of a structured *manuscript description* within a `<msDesc>` element defined for this purpose by the *TEI Guidelines* (TEI Consortium 2014: 295–337). Because of the extent and complexity of the manuscript description, its structure and content are described separately under subsection 11.1.6 below.

### 11.1.3 Encoding description

This part of the TEI header contains a concise summary of the transcription, encoding and annotation principles used in the edition, along with structured technical declarations referred to from the edition itself. The encoding description consists of five components: the project description, the sampling declaration, the editorial declaration, the tagging declaration and the character declaration. The *project description*, represented by a `<projectDesc>` element, contains prose paragraphs describing “the aim or purpose for which an electronic file was encoded, together with any other relevant information concerning the process by which it was assembled or collected” (TEI Consortium 2014: 36). In this case the project description contains a reference to this thesis and a link to the *Digital Editions for Corpus Linguistic* project. Since the present edition contains the whole text of the recipe collection, the *sampling declaration*, represented by a `<samplingDecl>` element, is used merely to state this fact and to briefly outline the different parts of the text included in the edition. In the case of collections contained in miscellany manuscripts (MSS Ad and C) containing multiple texts, a mention of the omission of these is also made.

The *editorial declaration*, represented by an `<editorialDecl>` element, is the largest of the encoding description components containing descriptive prose. It consists of a brief prose summary of the editorial principles of the edition, a reference to this thesis for more information, and several subcomponents detailing the editorial approach to individual aspects of the text in the form of prose paragraphs:

<b><code>&lt;correction&gt;</code></b>	“states how and under what circumstances corrections have been made in the text” (TEI Consortium 2014: 37);
<b><code>&lt;normalization&gt;</code></b>	“indicates the extent of normalization or regularization of the original source carried out in converting it to electronic form” (37);
<b><code>&lt;segmentation&gt;</code></b>	“describes the principles according to which the text has been segmented” (38), in the case of this edition, into text-structural and word-level units;
<b><code>&lt;stdVals&gt;</code></b>	(standard values) “specifies the format used when standardized date or number values are supplied” (38);
<b><code>&lt;interpretation&gt;</code></b>	“describes the scope of any analytic or interpretive information added to the text in addition to the transcription” (38).

The *tagging declaration* is the first of the structural declarations in the encoding description and contains: 1) descriptions of the different visual *renditions* that are

used by the annotation to describe the visual appearance of textual units in the original manuscript, and 2) a list of the number of occurrences of all the TEI XML elements used in the text itself (excluding the TEI header). The first of these is encoded as a series of <rendition> elements, each containing an @xml:id and a free prose description of the rendition in question (e.g. “Characters that have been written as superscript” for an @xml:id value of ‘sup’).<sup>11</sup>

The tag descriptions are represented as <tagUsage> elements contained within a <namespace> element, whose @name attribute defines the namespace to which the elements defined within it belong, namely ‘http://www.tei-c.org/ns/1.0’.<sup>12</sup> Each <tagUsage> element has two attributes: @gi, which contains the name (or *generic identifier*) of the element, and @occurs, which contains the total number of occurrences “within the outermost <text> element associated with the <teiHeader>” (TEI Consortium 2014: 43) for that element.

The *character declaration*, represented by a <charDecl> element contains a listing of all the non-alphabetic special characters used in the edition. Each special character is represented by a <char> element identified with an @xml:id attribute which is used as the value of the @ref attribute of <g> elements standing for the character in the transcription (see subsection 11.5.3). The properties of the character are described by two child elements of the <char> element. The <charName> (character name) “contains the name of a character, expressed following Unicode conventions” (TEI Consortium 2014: 183) (e.g. COMBINING ZIGZAG ABOVE). For characters that have a suitable representation in the Unicode or Medieval Unicode Font Initiative (MUFI) standards, the name used in these resources is used, and for others, one has been created following the same conventions. The <mapping> element contains a code point of the most suitable Unicode, MUFI or ad-hoc glyph that is recommended as the graphic representation of the special symbol. The @type attribute is used to characterise the relationship of the suggested representation to the symbol itself, using the value ‘exact’ for a code point that represents a semantically appropriate Unicode representation of the symbol in question, ‘approx’ for a code point that represents a graphical approximation of the symbol which is not semantically appropriate, and ‘PUA’ for a Private Use Area (PUA) code point which is not defined in the Unicode standard but has been defined either by the MUFI or by the present editor for representing the symbol in question.<sup>13</sup>

#### 11.1.4 Text profile

This section of the TEI header “provides a detailed description of non-bibliographic aspects of a text” (TEI Consortium 2014: 49). For the present edition, this consists of an indication of the likely creation date and place of the manuscript text, a listing (and rough proportions) of the languages used in it, a subject classification

<sup>11</sup> The decision to use free prose descriptions rather than formal descriptions using a stylesheet language such as CSS (Bos et al. 2011) or XSL-FO (Pawson 2012) was made because the formal interpretation of some of the renditions (see subsection 11.3.5) are dependent on the values of other attributes, namely @place and @rend.

<sup>12</sup> The modifications made to the TEI schema for the purposes of the present edition only involve attributes, which means that all of the elements used in the edition belong to the TEI namespace.

<sup>13</sup> For a more detailed description of the special characters used in the edition and their suggested representations, see subsection 11.5.3.

of the text, and its characterisation in terms of selected ‘situational parameters’. This metadata is not intended to reflect the results of a detailed analysis of the text, but to provide criteria for its classification in library databases or text corpora.

Information regarding the creation of the manuscript is encoded by including a `<creation>` element containing a `<date>` element and a `<placeName>` element as the first child of the `<profileDesc>`. Since the dating of the manuscript indicated here is intended for filtering texts in corpora, it has been left purposefully vague to maintain consistency, being indicated as a 50-year period (see subsection 11.1.1 above) even in cases where there is internal evidence that would seem to rule out a part of the indicated period. This period is indicated using the `@notBefore` and `@notAfter` attributes on the `<date>` element, which also contains a prose description of the period (i.e. “Late 15th century”). Since none of the manuscripts have any internal evidence of their region of origin, the place of origin for all of the MSS edited here is simply *England*, given as the content of the `<placeName>` element.<sup>14</sup>

The languages used in the edited text are encoded by a `<langUsage>` element which contains one `<language>` element for each language used in the text, with an `@ident` attribute identifying the language using an ISO 639-3 language code (SIL International 2011) and a `@usage` attribute indicating the percentage of the text that this language represents.<sup>15</sup> This quantification is not based on a detailed linguistic analysis of the text (represented as a separate non-editorial annotation overlay described in subsection 11.9.1) but on the editorial annotation of the principal language of each textual component (see 11.5.1) and the number of words contained within them, and is therefore merely a rough approximation.<sup>16</sup>

The subject classification of the edited texts is based on the Library of Congress subject, name and genre keywords (`<http://id.loc.gov/>`), and is encoded using a `<textClass>` element containing a series of `<keyword>` and `<term>` elements referencing the appropriate LC keywords describing the edited text. In addition to the keywords, a subject class code is also provided both in the LCC classification and the Dewey Decimal System, encoded using the `<classCode>` element with an appropriate `@scheme` attribute value.<sup>17</sup> The communicative characteristics of the edited text are also characterised by a ‘text description’ (`<textDesc>`), intended for the description of linguistic corpus texts. This element contains eight specialised child elements describing the text as a communicative event in terms of its communicative *channel*, *constitution*, originality or *derivation*, *domain*, *factuality*, extent of *interaction*, *preparedness*, and *purpose*, using both `@type` attribute values suggested in the *TEI Guidelines* (TEI Consortium 2014: 479-81) and brief prose descriptions.

<sup>14</sup> No formal value, such as an ISO country code is given, since the country codes are based on modern-day nation states and are thus potentially misleading.

<sup>15</sup> Since the proportions are expressed as percentages, languages with very minor presence (`< 0.5do` occur in the text).

<sup>16</sup> Although, as was pointed out in *Treatment of foreign words* in subsection 1, the ambiguity inherent in distinguishing between Middle English and ‘foreign’ words means that any such quantification, even when based on a detailed analysis of the language, is highly uncertain.

<sup>17</sup> As this data is the same for all six *PD* versions, it is not very useful in the context of the present edition, but is included because of its usefulness in case of the potential future inclusion of the texts in a linguistic corpus or a digital library system.

### 11.1.5 Revision history

The <revisionDesc> element of the header serves as a record of the creation and revision history of the edition.<sup>18</sup> It consists of a series of <change> elements, each containing a series of attributes describing a state of the edition at a given point in time, and a prose description of one or more of editorial actions leading to that state.<sup>19</sup> The attributes used to describe the change are:

- @status** indicates the status of a document following the change documented, using the following values defined by the *TEI Guidelines* (TEI Consortium 2014: 734-5): 'draft', 'submitted', 'cleared', 'approved' and 'published' (referring to the eventual online publication of the edition).
- @who** contains a pointer to a <persName> element in the <fileDesc> element of the header, identifying the person responsible for the change.
- @when-iso** indicates the date of the change using the ISO 8601 standard form (yyyy-mm-dd).
- @n** indicates the number of the version resulting from the change; prepublication versions have a version number of less than one.

In a single-editor work such as the present edition, the revision history is intended mainly as a documentation of the work process, although it does not provide very detailed information.

### 11.1.6 Manuscript description

Although the manuscript description is technically only one of the subcomponents of the *file description* part of the TEI header, it has a rather complicated internal structure in itself and in the case of a complex manuscript can easily take up as much space as the rest of the TEI header combined. The general framework on which the headers used in the present edition are based is described in the *TEI Guidelines* (TEI Consortium 2014: 295–337). The manuscript descriptions included in the present edition consist of six separate component elements: 1) <msIdentifier>, 2) <head>, 3) <msContents>, 4) <physDesc>, 5) <history>, and 6) <additional>.<sup>20</sup> The first of these, the <msIdentifier>, contains the full holding library and shelfmark information for the MS, organised using the elements <country>, <settlement>, <institution>, <collection>, and <idno>, along with the *secundo folio* or the first few words of the second folio, contained within a <msName> element of @type 'sec\_fol'. The second element, <head> has no substructure but simply contains the descriptive title or heading given to the manuscript in the library catalogue. After these identifying components, the following three components describe the manuscript in terms of its content, its physical properties, and its origin,

<sup>18</sup> Unfortunately, the revision history of the transcription files for the present edition is relatively sparse for the period preceding its conversion from the initial shorthand transcription format into TEI XML.

<sup>19</sup> For reasons of economy, only relatively major and systematic changes in the base data file are recorded in the revision history, minor corrections and individual changes not having been recorded.

<sup>20</sup> Like the components of the parent <fileDesc> itself, these components differ significantly in their extent, the first two ones being much smaller than the rest.

provenance and known history, while last one contains miscellaneous administrative information about the manuscript and a bibliography of sources referred to in the description.<sup>21</sup>

### Manuscript contents

The `<msContents>` element contains a description of the intellectual content of the manuscript, organised as a series of *manuscript items* and sub-items, each represented by its own `<msItem>` element. In order to reflect the hierarchical structure of these manuscript items—e.g. a collection of recipes having multiple sub-collections as well as a main part—the structure described below for a `<msItem>` can be nested indefinitely, usually leading to some of its components being present on a single parent element and others being divided among child elements describing its sub-items. *XML Example 2* contains an example of the structure of a full manuscript item with all of the elements used to annotate its content (apart from the general metadata annotations described above in subsection 11.1.1).

Of the elements listed in *XML Example 2*, only the `<locus>`, indicating the location of the item in the manuscript, and the descriptive `<title>` are mandatory, allowing also very brief items (individual lines or couplets of verse, recipes, aphorisms, etc.) to be listed efficiently. It should be noted that the `<msItem>` can nest within itself at any point, and the sub-items can contain only those components which are unique to them, relegating any information shared by all sub-items of a particular item (such as bibliographical references) to the parent element without the need to repeat them under every sub-item.<sup>22</sup>

Of the attributes used for the various elements, the `@n` attribute is used on each `<msItem>` to explicate their position in the hierarchy, the main items being identified by Arabic numerals (and secondary and tertiary items by alphabets and Roman numerals, respectively). The `@corresp` attribute is used on those items that are transcribed in the `<text>` of the file, the value being a pointer to the relevant `<group>`, `<text>`, or `<div>` element. The `@from` and `@to` attributes are used to indicate the extent of the item in terms of the foliation of the MS, as per the *TEI*

<sup>21</sup> While the above-mentioned order of elements within the manuscript description is strictly defined by the *TEI Guidelines*, the description of composite manuscripts—like MSS Ad and C of *Potage Dyvers*—makes an exception to this. For describing such manuscripts, the guidelines define the `<msPart>` element, which is intended to contain “information about an originally distinct manuscript or part of a manuscript, now forming part of a composite manuscript” (TEI Consortium 2014: 336) and which has the same content model as the `<msDesc>` element itself. Since a part of the physical description and most of the recorded history of the manuscript (as well as the identifying information described above) pertain not to any specific part of the composite manuscript but to the whole in its current state, the most natural solution would be to divide only the part-specific information—mainly the manuscript contents—within these `<msPart>` elements and to present the rest of the information for the whole manuscript. Unfortunately, the content model of the `<msDesc>` demands the `<msPart>` elements to occur last after all other content of the `<msDesc>`, which means that instead of the shared `<history>` and `<additional>` information following the individual description of the parts, the initial part of the descriptions of these manuscripts will contain all the information pertaining to the whole manuscript, followed by a series of similarly ordered `<msPart>` elements containing information specific to each part of the manuscript.

<sup>22</sup> This means that in most cases, sub-items are inserted in between the descriptive `<note>` element and the `<filiation>` element of the parent, and contain sub-item-specific versions of only the elements preceding this point (with possibly a separate `<textLang>` element if the sub-item is in a different language), the filiation, language and bibliographic references being indicated jointly on the parent element.

---

**XML Example 2: Structure of the description of a manuscript item.**


---

```

<msItem n="N" corresp="#transcribed_item">
  <locus from="NNr" to="NNv">ff. NNr-NNv</locus>
  <author key="">[name of the author if known; element omitted if anonymous]</author>
  <title type="supplied">[original title or modern descriptive title]</title>
  <rubric xml:lang="xxx">[rubric or MS title]</rubric>
  <incipit xml:lang="xxx">[beginning of the item]</incipit>
  <quote xml:lang="xxx">[some characteristic or noteworthy phrase from the item]</quote>
  <explicit xml:lang="xxx">[end of the item]</explicit>
  <note>
    <p>[brief description of the item in terms of its contents and structure]</p>
  </note>
  <foliation>
    <p>[information about the relationship of the item to other surviving manuscript versions
      of the same item, including links to possible editions and other secondary literature]</p>
  </foliation>
  <textLang mainLang="xxx" otherLangs="yyy zzz">[brief prose characterisation of the
    languages used in the item]</textLang>
  <bibl type="additional_info">
    <ptr type="biblref" target="#XXX" n="nnn"></ptr>
    <ptr type="biblref" target="#YYY" n="nnn"></ptr>
    <ptr type="biblref" target="#ZZZ" n="nnn"></ptr>
  </bibl>
</msItem>

```

---

*Guidelines* (303-4).<sup>23</sup> The @type of the <title> is indicated using the value 'supplied' for editorially supplied descriptive titles and 'uniform' for the normalised titles of established works (e.g. *Secretum Secretorum*).

### Physical description

The <physDesc> element contains a description of the manuscript as a physical artefact and is divided into five main components, many of which are further subdivided into specialised subcomponents. The general structure of the physical description hierarchy is shown in *XML Example 3*, with the content of its components explained below.

Of the five main components, the first, <objectDesc> is further divided into two subcomponents, <supportDesc> and <layoutDesc>, of which the subcomponents of the former—<support>, <extent>, <foliation>, <collation>, and <condition>—describe different codicological aspects of the manuscript. Within these elements, the prose descriptions are enclosed within the <p> element, except for the <extent> element which contains its textual content directly, and the <collation> element which also contains a list with each <item> representing the internal makeup of a single quire using the traditional formula, as well as two specialised paragraph-level elements, <signatures> and <catchwords>. Within the <extent>, the physical measurements of the manuscript are annotated with the <measure>

---

<sup>23</sup> In the case of items included in the transcription, the values refer to the @n attribute values of the relevant <pb> (page break) elements in the transcription, and in the case of other items, to the foliation extrapolated from the values of the <pb> elements of the transcribed section. In cases where the reference is to a single page, the @to attribute is omitted. It should be noted that since the <pb> element is located at the beginning of each page, the reference point of the @to attribute should properly be understood to be not the <pb> element itself, but rather the textual content following the <pb> element indicated by the value but preceding the following one, i.e. the content of the page marked by that folio reference.

---

**XML Example 3: Structure of the physical description of a manuscript.**


---

```

<physDes>
  <objectDesc>
    <supportDesc>
      <support>
        <p>[description of the writing material]</p>
      </support>
      <extent>[description of the size and number of folia]</extent>
      <foliation>
        <p>[description of any original foliation]</p>
      </foliation>
      <collation>
        <p>[description of the collation of the MS]</p>
        <list>
          <item>[quire 1]</item>
          ...
          <item>[quire n]</item>
        </list>
        <signatures>[listing of quire signatures]</signatures>
        <catchwords>[listing of quire catchwords]</catchwords>
      </collation>
      <condition>
        <p>[description of damage in the MS]</p>
      </condition>
    </supportDesc>
    <layoutDesc>
      <layout columns="n" writtenLines="nn nn">
        <p>[description of the layout]</p>
      </layout>
    </layoutDesc>
  </objectDesc>
  <handDesc>
    <summary>[summary of the hands used in the manuscript]</summary>
    <handNote xml:id="hand_xxx" script="style formality" medium="ink-colour" scribe="scribal_name">
      <p>[description of a single hand]</p>
    </handNote>
  </handDesc>
  <decoDesc>
    <decoNote type="xxx">
      <p>[description of a single aspect of decoration (rubrication, initials, etc.)]</p>
    </decoNote>
  </decoDesc>
  <additions>
    <p>[description of later additions made to the manuscript]</p>
  </additions>
  <bindingDesc>
    <binding>
      <p>[description of the current binding of the manuscript]</p>
    </binding>
  </bindingDesc>
</physDes>

```

---



element for the number of folia, and with the specialised `<height>` and `<width>` elements for the height and width of the manuscript.<sup>24</sup> The `<layoutDesc>` which forms the second component of `<objectDesc>` contains a `<layout>` element defining the layout of the manuscript (or of a specific manuscript part, if located within an `<msPart>` element).<sup>25</sup> The attributes encoding the number of columns and lines on the page—as per the definition in TEI Consortium (2014: 1115) can both take either a single value or a pair of values separated by space, indicating the minimum and maximum number of columns or lines in variable layouts.

The second component of the physical description—`<handDesc>`—contains a `<summary>` of the occurrence of different scribal hands in the manuscript, as well as descriptions of all the scribal hands identified in the manuscript, encoded in the form of individual `<handNote>` elements. The internal structure of the `<handDesc>` element and the mechanism used to link the various hands to the appropriate portions of the textual content is described below in subsection 11.4.1. The third component, `<decoDesc>`, describes the decoration applied to the manuscript in a similar format, containing first a brief `<summary>` of the level of decoration present in the manuscript, followed by a series of `<decoNote>` elements, each describing a specific aspect of the decoration, indicated by the value of the `@type` attribute, which in the present edition takes the values 'rubrication' and 'initials'. Manuscripts that contain decorative figures, such as line fillers or other ornaments, also contain a special `<decoNote>` with a `@type` value of 'svg', containing Scalable Vector Graphics (SVG) representations of these graphical elements, enclosed within `<figure>` elements.<sup>26</sup>

The last two elements of the physical description of the manuscript are concerned with the physical consequences of its later transmission history. The first of them, `<additions>`, contains a series of prose paragraphs describing any additions, whether physical, textual or graphical, made to the manuscript after its initial production. The production process of the manuscript is here understood to cover not only its initial writing, but also any decoration that is considered to form a part of the original design. The presence of emendations to the text of the manuscript is also mentioned, since it is often difficult to tell whether it has been performed as a part of the original production process or by a later user of the manuscript. The `<bindingDesc>`, like `<layoutDesc>` above, contains a single `<binding>` element which describes the current binding of the manuscript in the form of prose paragraphs, also transcribing any text occurring on the cover.

<sup>24</sup> These elements also encode the measurements formally using the `@unit` and `@quantity` (or `@min` and `@max`) attributes. The formal annotation of the size of the manuscript page allows these measurements to be used for formatting the manuscript page for diplomatic presentations of its content.

<sup>25</sup> Although the guidelines allow for multiple layout items to describe changes in the number of columns etc., the present edition opts to describe the layout of each manuscript part with a single `<layout>` element for reasons of simplicity; using multiple layout elements associated with specific spans of pages can significantly complicate the use of this data for formatting the visual representation of the manuscript page and would offer little benefit, as the changes in layout are described in the prose description in any case.

<sup>26</sup> The annotation of these graphical elements in the text is described in subsection 11.5.2.

## History and provenance

The known history of the manuscript, based both on secondary sources and internal evidence (such as marks of ownership and other annotations), is recorded by the `<history>` element (TEI Consortium 2014: 331-3), which contains a single `<origin>` element followed by a series of `<provenance>` elements, and finally one `<acquisition>` element. All of these elements have the same internal structure, consisting of one or more paragraphs of prose text, and employ the attributes `@when`, `@from`, `@to`, `@notBefore`, and `@notAfter`, defined in the *TEI Guidelines* (725-6), to date the events described by the element. Together, these three elements outline the known history of the manuscript, `<origin>` containing “any descriptive or other information concerning the origin of a manuscript”, `<provenance>`—which can occur multiple times—containing similar information “concerning a single identifiable episode during the history of a manuscript or manuscript part, after its creation but before its acquisition”, and `<acquisition>` describing “the process by which a manuscript or manuscript part entered the holding institution” (TEI Consortium 2014: 331-2).

## Additional information

The final element of the manuscript description—`<additional>`—contains three types of ‘additional’ or miscellaneous information about the manuscript, represented by the elements `<adminInfo>`, `<surrogates>` and `<listBibl>`. The first of these is used in the present edition to indicate the sources on which the manuscript description itself is based and to provide information about access to the original manuscript. The former is presented as a prose paragraph contained within a `<source>` element, in turn contained within a `<recordHist>` element, and the latter similarly as a prose paragraph within an `<availability>` element. The `<surrogates>` element is used to provide information about any surrogate copies—such as microfilm or digital images—that are available of the manuscript. The final component is a bibliography containing the reference information for all primary and secondary sources cited in the manuscript description, presented as a `<listBibl>` element containing two subordinate `<listBibl>` elements, the first of which (`@type='sources'`) contains bibliographical entries to all secondary sources in the form of structured bibliographic entries (`<biblStruct>`<sup>27</sup>), while the second (`@type='manuscripts'`) contains the reference information for all the other *PD* manuscripts in the form of `<bibl>` elements containing a `<msIdentifier>` element.

## 11.2 Textual structure

The *textual structure* of the document is here understood as the logical and functional structure of the text, abstracted from the visual organization of the writing on the manuscript page, by which it is often signalled. As was explained in chapter 10, the annotation of textual structure forms the principal structural hierarchy of the XML base data file and uses the principle of hierarchical *nesting* to indicate the logical structure of the text (see section 5.7). Another important concept

<sup>27</sup> The internal structure of the `<biblStruct>` element is described in the *TEI Guidelines* (119-41)

closely associated with the concept of nesting is *inheritance*, which refers to the fact that in hierarchical markup languages such as XML, elements nested within another element—their *children*—can *inherit* properties like attribute values from their *parent* element.

In line with its nature as a *discourse colony*, each manuscript collection of recipes is considered to constitute a single composite *text*. As described in section 9.2, some of the collections—namely MSS Ad and C—form a part of a composite manuscript, being surrounded by other texts not included in the edition, but this is not here considered to alter their status as individual texts, and does not affect their structure in the edition.<sup>28</sup> This interpretation is annotated by enclosing each recipe collection, including any peritextual material (such as tables of contents or bills of fare), within a <text> element, which—according to the *TEI Guidelines*—“contains a single text of any kind, whether unitary or composite, for example a poem or drama, a collection of essays, a novel, a dictionary, or a corpus sample” (TEI Consortium 2014: 150).

The composite nature of this text is indicated by a @type attribute value of 'collection', while the manuscript version contained in it is identified by the @xml:id attribute, whose value is the relevant manuscript signum (and the part identifier in the case of MS C) with all whitespace omitted (e.g. Additional5467). Also the primary language of the manuscript text is indicated by an @xml:lang attribute using the three-letter code defined in the ISO standard 639-3 (SIL International 2011) for Middle English, namely 'enm'. Similarly, the primary hand responsible for the majority of the collection (i.e. the hand of the original scribe) is indicated by the @pd:hand attribute following the method described in subsection 11.4.1, which is one of the two attributes used in the present edition that do not belong to the TEI namespace and thus constitutes an extension to the *TEI Guidelines*.<sup>29</sup>

### 11.2.1 The front, back and body of a collection

The contents of each <text> element are divided into the <front>, the <back> and the <body>, the first and second of which are optional and not present in all of the collections. As defined by the *TEI Guidelines*, the front matter of the collection contains “any prefatory matter (headers, title page, prefaces, dedications, etc.) found at the start of a document, before the main body” (TEI Consortium 2014: 150). In the context of this edition this means the possible table of contents of the recipe collection and any bills of fare prepended to the recipe collection itself. Correspondingly, the <back> element encloses any such paratextual matter following the body in the original document. Since all the items that commonly occur in the <front>, can also occur in the <back>—as they do also in the different *PD* versions—the two elements have an identical content model, being made up of a mixture of *textual divisions* grouping together related textual content (see subsection 11.2.2) and individual *component-level elements* (see subsection 11.2.3).

<sup>28</sup> Although it does mean that the manuscript description of these texts does also contain information about parts of the manuscript not transcribed in the edition.

<sup>29</sup> While the *TEI Guidelines* also define a @hand attribute (TEI Consortium 2014: 725, 766-7, 1035, 1473), but it is currently allowed only in very few elements, which do not include the <text> element or any other of the *textual containers* defined below. A feature request has been submitted to the TEI for expanding the scope of the @hand attribute (<<https://sourceforge.net/p/tei/feature-requests/480/>>), and this situation is subject to change.

In accordance with the *discourse colony* nature of the recipe collection itself which makes up the body of the text, it is annotated as a *composite text*, defined by the *TEI Guidelines* as “consisting of several components which are in some important sense independent of each other” (TEI Consortium 2014: 150), using the `<group>` element instead of the `<body>` element normally used for encoding the body of unitary texts. As defined in the *TEI Guidelines*, the `<group>` element contains a series of `<text>` elements making up the group, along with milestone elements that represent any page, column or line breaks occurring in between these component texts (see section 11.3 and subsection 11.4.3). This `<group>` element is further characterised by a `@type` attribute with the value 'collection' and an `<xml:id>` value of 'potage\_dyvers', which—as was mentioned in chapter 9—is the more or less arbitrary ‘title’ adopted for the family of collections edited here. For those versions of the *PD* family that consist of several separately titled subcollections or groups of recipes, namely MSS H279 and Ad, this substructure is annotated by enclosing these groupings within their own, uniquely identified, `<group>` elements of `@type 'recipegroup'`, located within the `<group>` representing the entire collection.

### 11.2.2 Recipe texts and textual divisions

In accordance with the characterisation of the recipe collection as a discourse colony, the individual recipes making up the body of a recipe collection are seen as individual texts in their own right, and are therefore tagged using the `<text>` element with a `@type` of 'recipe'. Each recipe of these `<text>` elements is uniquely identified using the `@xml:id` attribute, whose value consists of the `@xml:id` of the manuscript version suffixed with '\_rNN' where *NN* is the number of the recipe in the collection.<sup>30</sup> Each recipe `<text>` is structured similarly to the whole collection, with the possible recipe title and any notes pertaining to the whole of the recipe considered as a part of its *front* and the recipe text itself forming its *body*.

The contents of the `<front>` and `<back>` of the collection, on the other hand, are subdivided into *textual divisions* using the element `<div>`. As defined by the *TEI Guidelines* (TEI Consortium 2014: 152), the `<div>` element is used to represent a whole variety of textual subdivisions like books, parts, sections, chapters, acts, scenes and letters, depending on the genre of text. In the present edition, the `@type` attribute is used to indicate the type of division with the value 'contents' indicating a table of contents for the recipe collection or a part of it, the value 'menugroup' indicating a separately titled collections of menus or bills of fare, and the value 'menu' indicating an individual menu or bill of fare for a single occasion. All of these textual divisions, regardless of their type, have also been uniquely identified using the `@xml:id` attribute. *XML Example 4* provides an example of the basic document structure of a simple recipe collection.

It should be noted that the division of the front and back into textual divisions and of the body into recipe texts is not necessarily exhaustive, i.e. the `<div>` and `<text>` elements do not form a tessellated structure even in terms of the textual structure of the manuscript, as headings and other components (see subsection

<sup>30</sup> This numbering follows the actual order and number of recipes in the collection, including any partial recipes, which means that it does not always coincide with the numbers assigned to the recipes in the collection.

11.2.3) applying not to any textual division but to the whole of the collection (or its <front> or <back>) can be nested directly within the relevant element without an intervening <text> or <div>.

---

**XML Example 4:** *The basic XML document structure reflecting the textual structure of a recipe collection with no subcollections, a table of contents prepended and two bills of fare appended to the end.*

---

```
<TEI>
  <teiHeader>
    <!-- The metadata header. -->
  </teiHeader>
  <text type="collection" xml:id="ManuscriptID" hand="#primary_hand" xml:lang="enm">
    <front>
      <!-- Incipit for the recipe collection. -->
      <div type="contents" xml:id="contents_1">
        <!-- The table of contents. -->
      </div>
    </front>
    <group>
      <text type="recipe" xml:id="ManuscriptID_r1">
        <front>
          <!-- Heading(s) for the second recipe. -->
        </front>
        <body>
          <!-- Body text of the first recipe. -->
        </body>
      </text>
      <text type="recipe" xml:id="ManuscriptID_r2">
        <front>
          <!-- Heading(s) for the second recipe. -->
        </front>
        <body>
          <!-- Body text of the second recipe. -->
        </body>
      </text>
      ...
    </group>
    <back>
      <div type="menu" xml:id="menu_1">
        <!-- The first bill of fare. -->
      </div>
      <div type="menu" xml:id="menu_2">
        <!-- The second bill of fare. -->
      </div>
    </back>
  </text>
</TEI>
```

---

### 11.2.3 Headings, paragraphs, lists and items

The various kinds of textual divisions described above are made up of a series of lower-level structural items that the *TEI Guidelines* call “*component-level elements*” (TEI Consortium 2014: 152). The components used to structure the main flow of text are: <head> and <trailer> (headings or incipits, and explicits), <p> (paragraphs), <list> (lists) and <item> (items within lists). Of these, the <list> element is hierarchically distinct from the rest, since it does not contain text but rather groups together other components, namely *items* and *headings*, and is thus located on an intermediate level between the main textual divisions and textual components. A further intermediate layer between these textual components and

individual word units is added by the `<label>` element which is used within list items to annotate numerical labels assigned to the items. In addition to these textual components belonging to the main textual flow, there is also a parallel group of component level elements that are used to annotate textual content that lies outside the main textual flow, such as notes and comments added to the text either by the original scribe or a later annotator, folio numbers or quire signatures. These elements are described separately under subsection 11.4.3.

In this edition, `<list>` and `<item>` are used exclusively for the tables of contents and bills of fare found in the `<front>` and `<back>` of the texts. Due to their nature resembling textual divisions, all `<list>` elements have also been uniquely identified using the `@xml:id` attribute. In order to allow for example the linking of table of contents items to the corresponding recipes, all `<item>` elements have also been provided with `@xml:id` attribute values. In addition to these basic elements, the `<head>` element is used within lists to annotate the headings of lists (e.g. “first course”), and the `<label>` element is used within `<item>` elements to annotate numerical labels assigned to list items.

Annotated using these elements, the content of a recipe usually consists of one or more `<head>` elements within the `<front>` element of the recipe `<text>`, representing any recipe titles<sup>31</sup> and a single `<p>` element within the `<body>` for the running text of the recipe.<sup>32</sup> While these elements represent abstract textual components and not the physical layout of the manuscript as such, the documentary orientation of the present edition means that only textual components that are also visually signalled are annotated. This means that unless there is a clear indication of a break in the textual structure (a visual clue like empty line or indentation coinciding with a break in the information structure), a recipe is considered to form a single paragraph. Any original annotations (see subsection 11.4.3) whose point of reference is the whole of the recipe (such as reference numbers or other reader’s marks referencing the recipe) are annotated by a `<note>` element located in the `<front>` of the recipe text, while any notes referring to a certain point within the recipe have been annotated by a `<note>` located at that point within the `<body>` of the recipe. The actual physical location of the note (if not at its point of reference) has been indicated using the mechanism described below in subsection 11.3.5.

It should be noted that the `<head>` element (and its posterior counterpart, `<trailer>`) can occur on several different levels of the textual hierarchy, reflecting different levels of headings. The referential scope of the `<head>` or `<trailer>` elements should be interpreted to be its closest *ancestor* `<group>`, `<text>`, `<div>` or `<list>` element. I.e. a `<head>` element located inside a `<front>` element is considered to represent the heading of the whole `<text>`, while a heading nested inside a `<list>` represents the heading of that particular list.

<sup>31</sup> Some recipes have no title at all, while some have both a separate chapter number and a name, which have been annotated using two separate `<head>` elements. In the case of recipe numbers, the distinction between a heading and a note is made based on whether they are a part of the ‘original text’ or not. This does not necessarily mean that they are written in the hand of the original scribe—they can just as well be executed by the rubricator (who might or might not be the same person as the original scribe), but it does mean that they were *planned* by him. This might be indicated for example by traces of instructional notes to the rubricator (which—as far as they survive—would themselves be annotated using the `<note>` element) or by spaces left for the heading.

<sup>32</sup> The number of paragraphs within a recipe is not limited to one, but no recipe in the present edition contains more than one paragraph.

### 11.2.4 Arbitrary text segments

An *arbitrary text segment* is here taken to mean any span of text below the component level which does not correspond to any unit of textual structure and needs to be annotated for some feature not covered by any of the specialized elements defined by the *TEI Guidelines*. In this edition these features include:

- 1) displaced segments of text (see subsection 11.3.5), and
- 2) segments of the original text written in a different hand, usually by the rubricator (see subsection 11.4.1).

These arbitrary segments of text—which can range from a single character to the greater part of a paragraph—are annotated using the generic element `<seg>` defined by the *TEI Guidelines* for this purpose (TEI Consortium 2014: 508-12). This element implies no semantic properties, simply serving as a carrier for the attributes that apply to the segment in question. In terms of attribute inheritance, the `<seg>` element serves as a textual container, as described under subsection 11.2.6 below.

### 11.2.5 Words, numbers, characters and punctuation

The lowest level of text-structural annotation in the edition is the annotation of individual word-level units. Although this level of annotation is no longer strictly text-structural, but rather syntactic or morphological, it nevertheless forms the endpoint of the structural hierarchy and is thus best discussed here. Since this word-level annotation is also the principal means of indexing or addressing the text, it covers the whole of the textual content in a tessellated fashion, i.e. with no exclusions or overlaps. This means that all significant whitespace characters—i.e. ones that represent space present in the original document—are also included within the preceding word-level element.<sup>33</sup>

Each word—as defined in *Word-units as textual coordinates* in subsection 10.1.3—is annotated using the `<w>` element (TEI Consortium 2014: 542), while numbers—whether Roman or Arabic—are annotated using the `<num>` element (90), punctuation using the `<pc>` element (547) and individual characters occurring outside words using the `<c>` element (547).<sup>34</sup> In order to isolate the lexical item represented by the `<w>` element from any surrounding space and to explicate its transcriptional status, the transcribed content of each `<w>` element—excluding any word-final whitespace—is placed within the `<orig>` element, defined by the *TEI Guidelines* for annotating original documentary readings (as opposed to editorially normalised ones) (TEI Consortium 2014: 81). The use of the `<w>` and `<orig>` elements and the treatment of original whitespace is demonstrated in *XML Example 5*. In order to allow their automated analysis, the numerical value of all

<sup>33</sup> This means that transformations and applications used to manipulate the edition should preserve whitespace within these elements and disregard it outside of them.

<sup>34</sup> The decision between the `<c>` and `<pc>` elements is not based on the character contained within them as such, but rather on its textual function. This means that the same symbols can occur both within a `<pc>` and a `<c>` element, depending on whether they function as a punctuation mark or a standalone character. The punctuation characters (usually the punctus) that are frequently used to separate a number from the surrounding text (Hector 1966: 41) are considered to be a part of the number and included within the `<num>` element.

numbers is annotated using the @value attribute on the <num> element.<sup>35</sup> Consecutive punctuation characters that are interpreted to fulfil a single function (e.g. a series of *puncti* and *virgulae* signalling the end of a recipe) are annotated as a single punctuation character.

---

**XML Example 5:** *Representation of the original form of a word in the transcription.*

---

```
<w xml:id="w2523"><orig>bake</orig> </w>
```

---

In order to create the *textual coordinate system* defined in subsection 10.1.3, each of these word-level elements—which cover the entire textual content of the edition—is provided by a unique identifier as an @xml:id attribute value, allowing them to be explicitly referenced and annotated by stand-off *annotation overlays*. For words which have been split into two by a line break (see below), the same base identifier is used, suffixed by the letter *a* for the initial part and *b* for the final part (e.g. 'w1254a' and 'w1254b').

### 11.2.6 Textual containers

In terms of representation of textual content, there are fundamentally two kinds of elements in the TEI XML specification. Some elements—including all of the elements used to annotate textual structure, along with some others like <add>, <note> and <fw>—are ontologically *textual containers* in the sense that they introduce and describe a segment of original text. Other elements like <hi>, <del> and <ref>—although often structurally similar or even parallel to the aforementioned elements—do not introduce a new text segment—even if they contain one—but rather describe a feature or an event affecting a pre-existing text segment or relating to it. This conceptual distinction, which is not explicitly made in the *TEI Guidelines*, is not merely theoretical but also has some important practical consequences.

The logical structure described above means that in TEI XML documents, attributes have a variable scope depending on their parent element. This, in turn, has practical implications in terms of the inheritance of those attributes that can pertain not only to the element itself but also to its textual content. Of the attributes defined by the *TEI Guidelines* and used in this edition, these include @xml:lang, @rend and @rendition, as well as the attribute @pd:hand defined for the present edition as an extended replacement for the TEI attribute @hand. In textual containers these attributes characterize not only the element itself but also all of its textual content and descendant elements (unless they specify a new value for the attribute in question), but in others (such as <hi> and <del>) they describe only the phenomenon described by the element.

Since the word *alteration* in *XML Example 6* was written by the original scribe, it should inherit the value of the @hand attribute not from its parent element <del> (whose attribute only defines the hand in which the word was struck through) but from the <div> element. The word *correction*, however, was added by the annotator and should inherit it from the <add> element instead of the <div> element. The

---

<sup>35</sup> Ordinal numbers are also marked apart from cardinal ones by @type value of 'ordinal'.



XML Example 6: Attribute inheritance.

```
<div hand="#scribeA">
  <w><orig>This</orig> </w>
  <w><orig>sample</orig> </w>
  <w><orig>text</orig> </w>
  <w><orig>contains</orig> </w>
  <w><orig>an</orig> </w>
  <subst>
    <del type="marked" rend="singlestrike" hand="#annotatorA">
      <w><orig>alteration</orig> </w>
    </del>
    <add type="supralinear" hand="#annotatorA">
      <w><orig>correction</orig> </w>
    </add>
  </subst>
  <w><orig>by</orig> </w>
  <w><orig>a</orig> </w>
  <w><orig>later</orig> </w>
  <w><orig>annotator</orig></w>
</div>.</pc>
```

difference between these structurally similar elements is that while `<del>` contains a segment of the original text, to which the event described by the element applies, `<add>` is a textual container and contains a new segment introduced into the text by the event (and the hand) described by the element. In order to represent this difference and accurately model the production processes of manuscript texts, the document structure of this edition assumes that elements inherit their attributes from their closest textual container ancestor. The category of textual containers is in this edition made up of the following elements: `<add>`, `<back>`, `<body>`, `<c>`, `<div>`, `<front>`, `<fw>`, `<group>`, `<head>`, `<item>`, `<label>`, `<list>`, `<note>`, `<num>`, `<p>`, `<pc>`, `<seg>`, `<text>`, `<trailer>`, and `<w>`.

Incomplete or divided textual containers

Cases where textual container elements have been left incomplete or are divided into several parts by manuscript phenomena represented by other elements are indicated by the `@part` attribute, defined by the *TEI Guidelines*.<sup>36</sup> It should be noted that for the purposes of facilitating processing of the text, this mechanism is also used for textual containers split by milestone elements like `<pb/>`, `<cb/>` and `<lb/>`, even though this is not necessitated by the XML data model. This applies mainly to `<w>` (word) elements, which are split at line, column and page breaks and `<p>` (paragraph) elements, which are split at page breaks.<sup>37</sup> The following values of the attribute, defined by the *TEI Guidelines*, are used for this purpose:

@part	(on textual container elements)
'Y'	(yes) the division is incomplete
'I'	(initial) the initial part of a divided element
'M'	(medial) a medial part of a divided element

<sup>36</sup> Elements that are incomplete due to damage are not annotated using the `@part` attribute but instead contain a `<gap>` element representing the missing content (see subsection 11.6.3).

<sup>37</sup> Containers that do not directly contain text (such as `<list>` or `<div>`) are not fragmented over line, column or page breaks.

---

'F' (final) the final part of a divided element

---

The values 'I', 'M' and 'F' are used only when all of the parts of the container are present but separated from each other by some other element, allowing for their reconstitution into a single element. Whenever some part of the container is missing from the original, the value 'Y' is used.<sup>38</sup> However, it should be noted that this attribute is *only* used on those textual containers which exclude some of their original textual content, meaning that it is *not* used on those container elements which contain a <gap> element representing the lost content (see subsection 11.6.3).<sup>39</sup> In terms of the identifiers assigned to textual containers using the @xml:id attribute, partial containers (part='Y') are identified just as complete ones, while the initial, medial and final parts of a container divided into multiple parts ((part='I', 'M' or 'F'), all receive the same base identifier suffixed by the letter *a* for the initial part, *b* for the second part and so on (e.g. 'p123a' (I), 'p123b' (M) and 'p123c' (F)).

### 11.2.7 Linguistic identity of textual items

Although the linguistic analysis of the text does not belong to the tasks of the editor, the analytical annotation of textual structure is here considered to include the indication of the linguistic identity of textual items down to the component level. This means that the principal language of textual divisions down to the level of headings, paragraphs, lists, etc. is annotated to provide a general indication of the linguistic makeup of the text, but their internal linguistic composition is left to a separate linguistic analysis of the textual content.<sup>40</sup> The language represented by the various component-level items is indicated by the @xml:lang attribute defined by the *TEI Guidelines* as a global attribute available for all elements, the value of which is inherited by elements based on the principle of *attribute inheritance* described above (TEI Consortium 2014: 740).

In practice this means that the principal language of each recipe collection (Middle English) is indicated by the @xml:lang value of the main <text> (of @type 'collection'), and structural units down to the level of paragraphs, whose predominant—or *matrix*—language differs from this, are indicated by the appropriate @xml:lang value.<sup>41</sup> The values of the @xml:lang attribute follow the ISO 639-3 standard, which also provides codes for “extinct, ancient, historic, and constructed languages” (SIL International 2011), including those found in the texts edited here, namely Middle English ('enm'), Anglo-Norman French ('xno') and Latin ('lat'). The

<sup>38</sup> The value 'N' (no), also defined by the TEI Guidelines, is not used as it is considered the default value, indicated by the absence of the @part attribute.

<sup>39</sup> This means that a paragraph which has lost a part of its contents with a lost folio is annotated with a @part attribute, as the paragraph element ends at the end of the page preceding the missing one and the <gap> element representing the lost content is located outside of it, whereas a word which has lost some letters with the excision of the edge of the page (or the paragraph containing it) is *not* annotated with a <part> attribute, as the missing content is represented by a <gap> element within the <w> element.

<sup>40</sup> The detailed linguistic analysis of the textual content, for which this structural annotation of language forms a context, is described in subsection 11.9.1 below.

<sup>41</sup> The fact that the annotation of language is here limited to the component level means that the <foreign> element, which is defined by the *TEI Guidelines* for indicating foreign words (TEI Consortium 2014: 68), is not used in this edition.

value 'eng' is used for any modern English annotations and the value 'und' (undetermined) for any components whose language cannot be determined.

## 11.3 Physical structure

Unlike in more display-oriented markup languages like HTML, where the annotation of textual structure has implications for the presentation or layout of a document, the annotation used in this edition makes a clear-cut distinction between the logical textual structure and the presentation or layout of the original document and uses separate methods to annotate them. Thus—unlike in for example HTML—things like the beginning of a new textual paragraph or the introduction of a heading is *not* considered to have any implications for the layout of the page, such as the beginning of a new line. Instead, all features of the original layout, such as line and page changes, and any deviations from the regular left-to-right, top-to-bottom flow of the text are explicitly annotated. This section describes the annotation used to indicate the layout of the document, including division into pages, columns and lines, as well as the location of any textual elements displaced from their normal place in the flow of text.

Since this edition is principally concerned with the representation of unique texts and the artefacts containing them instead of the abstract work, the reference system for describing the structure of the manuscript—based on division into pages, columns and lines—should be considered equally important as the description of textual structure. Since these divisions are likely to overlap with the textual divisions described above, they are annotated not by enclosing elements, but by empty *milestone elements* placed at the beginning of the division annotated by the element. Due to the aforementioned separation of the textual and physical structure of the manuscript, the milestone elements used to annotate the physical structure and layout of the manuscript should not be considered a part of the hierarchical nesting structure, i.e. they should not be considered to 'belong' or be associated with their parent elements.

As Pierazzo (2011: 470) has observed, handwriting is always more or less irregular in its spacing and dimensions, and therefore the spatial dimensions of the manuscript page are difficult to represent in the digital medium. The approach taken in the present edition is similar to the one taken in the *Jane Austen's Fiction Manuscripts Digital Edition*,<sup>42</sup> described by Pierazzo (2011), in that it represents the position of visual phenomena on the page "in an approximate and relative way" instead of using absolute measurements (Pierazzo 2011: 470). This decision is based on the argument that a representation of the spatial aspects of the page in terms of absolute measurements could be seen to correspond to the graphetic—or even graphic—representation of the textual content, while a suitably categorised schematic representation of the spatial relationships between the textual elements on the page can be seen to represent the same level of abstraction as the graphemic transcription of the textual content (Gabler 2007: 204). As was argued in subsection 5.4.1, the reduction of detail involved in the abstraction of analogue data into a suitable number of discrete categories actually provides the representation with

<sup>42</sup> <<http://www.janeausten.ac.uk/>>

more analytical power rather than less. In the present edition, the level of granularity used for the encoding of manuscript layout has been chosen to allow both the consistent encoding and analysis of the layout of the manuscript page using a reasonable number of analytical categories, and its reasonably accurate visual representation.

### 11.3.1 Foliation

Although Pierazzo and Stokes (2010: 422) have pointed out that the page is not always the most fundamental unit of documentary structure, there are fortunately very few visual elements in the *Potage Dyvers* that cross the limits of an individual page, allowing us to treat it as such in the present edition. The division of the manuscript into *pages* and *folia* is achieved by inserting the specialised milestone element `<pb>` (TEI Consortium 2014: 114) before every page of the manuscript text, including the first one. The editorially established reference foliation is annotated using the `@n` attribute on the `<pb>` element. The format used for the foliation is '[no. of folio][r/v]' (recto/verso).<sup>43</sup> Since all of the manuscripts included in this edition have been foliated at some point of their history, it is convenient to use the manuscript foliation as the basis for the encoded foliation. However, since the main purpose of the foliation is to provide an unambiguous way for referring to the pages of the manuscript, the editorial foliation reflects the present physical state of the manuscript, numbering each folio in sequence from the beginning of the manuscript, even if this means deviating from an established foliation or signature scheme.

Leaves that have been lost through damage and are thus no longer part of the manuscript are not indicated by a `<pb>` element, but rather by a `<damage>` and a `<gap>` element with attributes indicating the number of physical folia and pages of text that have been lost (see section 11.6), placed within a `<note>` element of `@type 'damage'` located between the outermost incomplete element (with a `@part` attribute) of the preceding page and the `<pb>` element of the following page, as demonstrated in *XML Example 7*.<sup>44</sup>

Another special case are extra folia at the beginning of a manuscript that have textual content but have not been included in the canonical foliation (e.g. added leaves with later tables of contents etc.). This situation is found in MS As, which has a series of post-medieval leaves surrounding the original manuscript, of which the last prepended leaf has annotations in two 17<sup>th</sup>-century hands. This leaf has been accommodated by identifying it as folio 0.

### 11.3.2 Columns

On pages with multiple columns, the division into columns is annotated using the specialized milestone element `<cb>` (TEI Consortium 2014: 114). Division into

<sup>43</sup> The `<pb>` element is always placed as high as possible in the document hierarchy. Namely, when a page break co-occurs with a break between two structural elements, the `<pb>` element is placed between them as their sibling element.

<sup>44</sup> A textual note providing further information about the loss, such as any traces of the lost pages or peculiarities in the collation of the manuscript at the damaged point, may be linked either to the `<gap>` element or the `<damage>` element containing it, depending on whether the note refers primarily to the lost folia or to the damage that caused their loss.

---

XML Example 7: Annotation of missing folia.

---

```

<text type="recipe" part="Y">
  <front>
    <head>Recipe title</head>
  </front>
  <body part="Y">
    <p part="Y">
      incomplete recipe text paragraph
    </p>
  </body>
</text>
<note type="damage">
  <damage agent="loss" degree="1.0" extent="a single folio" group="15" resp="#VM">
    <gap reason="damage" quantity="2" unit="pages" resp="#VM"/>
  </damage>
</note>
<pb n="24r"/>
<text type="recipe" part="Y">
  <body part="Y">
    <p part="Y">
      incomplete recipe text paragraph
    </p>
  </body>
</text>

```

---

columns is done on a per-page basis and the columns are named in an alphabetical sequence horizontally from left to right (i.e. the first column is *a*, the second *b*, etc.) and the name is entered as the value of the @*n* attribute of the <cb> element. This means that a page break element (<pb>) automatically resets the column layout to the default single-column format.<sup>45</sup>

In terms of encoding, manuscript pages with multiple different column layouts are problematic. In order to accommodate these kinds of pages, all separate column blocks (including the first one) on a multi-column page are marked with an initial <cb> element, with the value(s) of the @*n* attribute indicating their horizontal location and span (multiple values being separated with whitespace); i.e. the value for a two-column wide text block spanning the second and third column of a three-column page would be 'b c'. These column blocks are always contiguous—i.e. a half-page wide text block with a full-page width section in the middle is considered to form three separate column blocks—and they flow from the top of the page without overlap, the length of each column block being determined by the number of line breaks it contains (see below). *XML Example 8* illustrates the annotation of various types of text column configurations.

### 11.3.3 Lineation

In order to allow for both the accurate reconstruction of the manuscript page layout and the explicit identification of individual manuscript lines, all manuscript lines are explicitly indicated by preceding their content with the line break element <lb>. As was mentioned above, no other element—including 'block-level' elements like <p>, <div> and <item>—is considered to imply the beginning of a

---

<sup>45</sup> Like the page break (<pb>), the <cb> element is also placed as high as possible in the document hierarchy, i.e. when a column break co-occurs with a break between two structural elements, the <cb> element is placed between them as their sibling element.

XML Example 8: Annotation of multicolumn page layouts.

Text 1	Text 2
--------	--------

```
<pb xml:id="f.1r"/>
<cb n="a"/>
  Text 1
<cb n="b"/>
  Text 2
```

Text 1	Text 2
Text 3	
Text 4	Text 5

```
<pb xml:id="f.1r"/>
<cb n="a"/>
  Text 1
<cb n="b"/>
  Text 2
<cb n="a b"/>
  Text 3
<cb n="a"/>
  Text 4
<cb n="b"/>
  Text 5
```

Text 1	Text 2	Text 3
	Text 4	
	Text 5	Text 6

```
<pb xml:id="f.1r"/>
<cb n="a"/>
  Text 1
<cb n="b"/>
  Text 2
<cb n="c"/>
  Text 3
<cb n="b c"/>
  Text 4
<cb n="b"/>
  Text 5
<cb n="c"/>
  Text 6
```

Text 1	Text 4
Text 2	
Text 3	Text 5

```
<pb xml:id="f.1r"/>
<cb n="a"/>
  Text 1
<cb n="a b"/>
  Text 2
<cb n="a"/>
  Text 3
<cb n="b"/>
  Text 4
<cb n="b"/>
  Text 5
```

new line. The left-hand example in *XML Example 9* represents a heading, followed by two paragraphs of three lines, all separated from each other by an empty line, making up a total of nine lines. The right-hand one also represents a heading followed by two paragraphs, but here the first paragraph begins on the same line as the heading and the second one follows the first with no empty line in between, making up a total of six lines.

XML Example 9: Annotation of lineation.

```
<lb/><head>This is a heading </head>
<lb/>
<p>
  <lb/>Text in the first paragraph.
  <lb/>More text in the first paragraph.
  <lb/>Even more text in the first paragraph.
</p>
<lb/>
<p>
  <lb/>Text in the second paragraph.
  <lb/>More text in the second paragraph.
  <lb/>Even more text in the second paragraph.
</p>
```

```
<lb/><head>This is a heading </head>
<p>
  Text in the first paragraph.
  <lb/>More text in the first paragraph.
  <lb/>Even more text in the first paragraph.
</p>
<p>
  <lb/>Text in the second paragraph.
  <lb/>More text in the second paragraph.
  <lb/>Even more text in the second paragraph.
</p>
```

11.3.4 Empty space

Empty space, wider than a normal inter-word space and left on the writing line by the scribe for whatever reason is annotated using the `<space>` element (TEI Consortium 2014: 389). This element has no content, being an empty element placed in the text stream at the point of the empty space. The dimensions of the space are indicated using the `@dim`, `@unit` and `@quantity` attributes. Since the element

is in this edition used only for inline spaces, the value of the `@dim` attribute is always 'horizontal'. Of the options defined by the *TEI Guidelines*, the `@unit` used for recording spaces and other inline phenomena (see also subsection 11.6.3) in this edition is 'chars' (characters), which is here understood to correspond to an *n-unit* or the space of two minims. The `@quantity` attribute takes a numerical value, representing the length of the space in the units specified.

### 11.3.5 Displaced items

The concept of *displaced items* is used to refer to items—either simple spans of text like individual words or phrases, or structured textual elements like headings or even paragraphs—on the manuscript page whose physical location does not coincide with their logical point of reference in the textual stream. Displaced items include segments of the original text (<seg>, <head>, etc.), additions to the original text flow (see subsection 11.4.2), paratextual elements such as marginal notes referring to points or segments in the original text flow but not intended to be a part of it, and *forme work* (see subsection 11.4.3). While these items are transcribed at their logical point of reference, their physical location on the page is indicated separately using the methods described below.<sup>46</sup>

#### Placement of displaced items

Displaced elements are identified by the presence of the `@pd:place` attribute, whose value indicates the general location of the element on the manuscript page.<sup>47</sup> In order to enable the annotation of displacement for all *textual containers* that occur outside of the textual stream in the manuscripts, the scope of the `@pd:place` attribute has been expanded from that defined for the corresponding attribute in the *TEI Guidelines* (TEI Consortium 2014: 756-7).<sup>48</sup> For the purposes of indicating the placement of displaced items, the manuscript page has been divided into five principal *regions*, shown in *Figure 11.1*. In addition to these five principal regions which are defined in relation to the manuscript page, the present edition also defines three values for indicating the placement of displaced elements in relation to

<sup>46</sup> A similar approach has been used for additions and substitutions in *The Austen Digital Edition*, of which “every word has been isolated and moved according to its relative position using CSS-based coding” (Pierazzo 2011: 472), although Pierazzo does admit that “[t]he attempts have not been entirely successful and sometimes a less precise approximation than the one that was hoped for had to be accepted” (472). The present edition attempts to circumvent this problem and achieve a faithful representation of the spatial relationships between the various elements of the page by annotating the position of displaced items declaratively using a set of attribute values specifically designed for the purpose and converting this encoding to a suitable presentation format, such as CSS-formatted HTML only at a later stage.

<sup>47</sup> As indicated by the namespace prefix, this attribute is the second of the two attributes defined specifically for the present edition.

<sup>48</sup> The *TEI Guidelines* allow the `@place` attribute on elements belonging to the `add.placement` attribute class, namely <add>\*, <addSpan>, <figure>, <fw>\*, <label>\*, <metamark>, <notatedMusic>, <note>\*, and <witDetail>, of which only the ones marked with an asterisk are used in the present edition. This means that the schema of the present edition adds the elements <head> and <seg> to the `att.placement` class. While the `@place` attribute is actually only used on the elements <add>, <fw>, <head>, <label>, <note> and <seg>, all textual containers have been included for reasons of consistency. A feature request has also been submitted to the TEI for adding these elements to the `att.placement` class (<<http://sourceforge.net/p/tei/feature-requests/479/>>), which would allow the present edition to comply with the *TEI Guidelines* in this regard without modifications.

text *columns* on multicolumn pages and three values for indicating placement in relation to the logical point of reference. This means that there are a total of eleven values used for indicating the placement of items displaced from the textual flow:

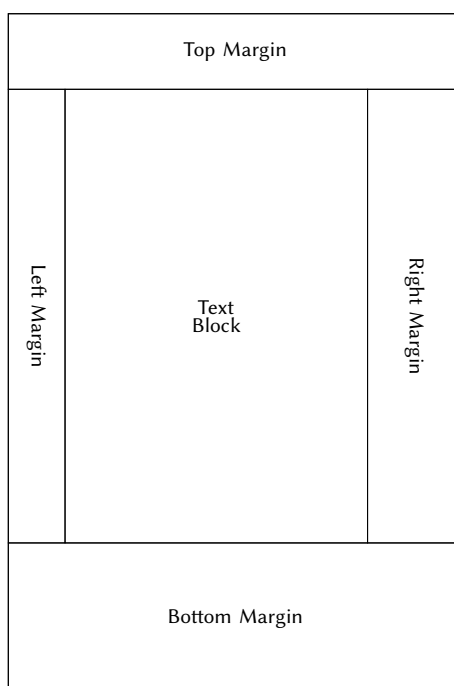
@place	(on <i>textual container</i> elements)
'text-block'	inside the text block
'margin-top'	in the top margin of the page
'margin-bottom'	in the bottom margin of the page
'margin-left'	in the left margin of the page
'margin-right'	in the right margin of the page
'column-block'	inside the text block of the current column on a page with multiple columns <sup>49</sup>
'columnmargin-left'	in the marginal space between the current column and the one to its left in a multicolumn layout
'columnmargin-right'	in the marginal space between the current column and the one to its right in a multicolumn layout
'above'	in the interlinear space directly above the point of reference
'aboveprev'	in the interlinear space directly above the word-unit preceding the point of reference
'abovenext'	in the interlinear space directly above the word-unit following the point of reference

The text block is the area of the page reserved for text, often (but not always) delimited by ruled lines, regardless of whether it has actually been filled with text or not. In the case of empty unruled pages, the measurements of the text block on other non-empty pages are used to define the area of the text block. The margins are the areas of the page outside the text block. In dividing the margins into regions, the top and bottom margins take precedence over the left and right margins, extending from the left edge of the page to the right, covering the areas above and below both the text block and the side margins. The left and right margins cover the remaining area to the left and right of the text block. The three 'columnar' values in the list are only applicable to manuscript pages divided into several columns (see subsection 11.3.2 above), and the 'columnmargin' values are naturally only applicable if there is a column on the appropriate side of the current column; the 'margin-left' and 'margin-right' values should be used for items displaced to the outside of the outermost columns.<sup>50</sup> The last three values represent items that have been displaced into the interlinear space above the point of reference but remain horizontally anchored to it, being placed either above the point of reference (usually between two words or two characters of a word in the

<sup>49</sup> 'Current' in this context refers to the column in which the logical point of reference for the displaced item is located.

<sup>50</sup> These values do overlap with the 'text-block' value, and there are cases where either could be used to the exact same effect. However, there are cases (e.g. marginal numbers in multicolumn lists) which are difficult to represent accurately with reference to the entire text-block, and the column-specific values are intended to provide a more semantically appropriate way of encoding these kinds of layouts. As a general rule, the 'columnmargin' values should be preferred in cases where there is a clear and regular margin (whether ruled or not) between the columns, and the 'text-block' value in cases, where the items between the columns can, in the absence of a clear margin, equally well be considered to lie in the middle of the entire text block.





**Figure 11.1:** *The five principal regions denoted by the @place attribute on a typical manuscript recto page.*

case of word-internal additions) or above the preceding or following word-unit (i.e. word, number, stand-alone character or punctuation symbol).

### Horizontal position

The placement of the item within the region defined by the @place attribute is indicated using the @rendition and @rend attributes on the relevant textual container element.<sup>51</sup> The difference between these two semantically similar attributes lies in the fact that while @rend takes a descriptive text token as its value, @rendition points to “a description of the rendering or presentation used for this element in the source text” (TEI Consortium 2014: 5), contained in the header of the document. The *horizontal alignment* of displaced items is indicated by the @rendition attribute in relation to the containing region. For this purpose, each region is divided horizontally into nine zones, shown at the top of *Figure 11.2*. Since the left and right margins (and the column margins) are considerably narrower than the other regions, a reduced version of the division, shown at the bottom of *Figure 11.2*, is used for items located within them.<sup>52</sup>

<sup>51</sup> In this edition, these global attributes are used for describing a variety of presentational aspects of items in the original document. For other uses of these attributes, see subsections 11.4.2 and 11.5.1.

<sup>52</sup> Since the region for the 'above' value is considered to be a point, it is considered to have only a centre zone. For the other interlinear values ('above\_prev' and 'above\_next'), the region is considered to be the width of the preceding or following word-unit and to be divided into the same five zones as the side margins.

left	left left centre	left centre	centre left centre	centre	centre right centre	right centre	right right centre	right
left	left centre	centre	right centre	right				

Figure 11.2: The horizontal zones within a page region denoted by the @rendition attribute.

In addition to defining the horizontal zone in which the displaced item is located, the value of the @rendition attribute also indicates the *anchor point* of the displaced item, the options being the *left end*, *right end* and *centre* of the item. The anchor point is relevant in the case of long items which extend over several zones or even several page divisions, as the attributes indicate the region and zone in which the anchor point of the item resides. For individual free-floating items, the centre of the item is the default choice, while the left and right end anchor points are used for encoding the items in indented or right-justified lists and other features that are aligned by their ends rather by their centre points, and adjacent displaced items that occupy the same zone but are slightly offset from each other horizontally.<sup>53</sup> The following renditions for the horizontal alignment and anchor point of a displaced item are defined in the header and can be referred to in the @rendition attribute:

@rendition	(on <i>textual container</i> elements)
'#c_al_l'	centre aligned to the left zone of the region
'#c_al_llc'	centre aligned to the left left centre zone of the region
'#c_al_lc'	centre aligned to the left centre zone of the region
'#c_al_clc'	centre aligned to the centre left centre zone of the region
'#c_al_c'	centre aligned to the centre zone of the region
'#c_al_crc'	centre aligned to the centre right centre zone of the region
'#c_al_rc'	centre aligned to the right centre zone of the region
'#c_al_rrc'	centre aligned to the right right zone of the region
'#c_al_r'	centre aligned to the right zone of the region
'#l_al_l'	left end aligned to the left zone of the region
'#l_al_llc'	left end aligned to the left left centre zone of the region
'#l_al_lc'	left end aligned to the left centre zone of the region
'#l_al_clc'	left end aligned to the centre left centre zone of the region
'#l_al_c'	left end aligned to the centre zone of the region
'#l_al_crc'	left end aligned to the centre right centre zone of the region
'#l_al_rc'	left end aligned to the right centre zone of the region

<sup>53</sup> It could be argued that regular structures such as lists, whose items consist of a number or other label and the item itself, both of which are horizontally aligned to a specific point, are quite different in nature to items that have been displaced to a different part on the page, but the justification used here for annotating both of them using the same structure is that they are both instances of the use of spatial positioning for text-structural or semantic purposes, which is in itself a useful analytical category.

'#l_al_rrc'	left end aligned to the right right zone of the region
'#l_al_r'	left end aligned to the right zone of the region
'#r_al_l'	right end aligned to the left zone of the region
'#r_al_llc'	right end aligned to the left left centre zone of the region
'#r_al_lc'	right end aligned to the left centre zone of the region
'#r_al_clc'	right end aligned to the centre left centre zone of the region
'#r_al_c'	right end aligned to the centre zone of the region
'#r_al_crc'	right end aligned to the centre right centre zone of the region
'#r_al_rc'	right end aligned to the right centre zone of the region
'#r_al_rrc'	right end aligned to the right right zone of the region
'#r_al_r'	right end aligned to the right zone of the region

Vertical offset and rotation

The *vertical position* of the item is expressed as an offset from a reference point and measured in lines of text. In order to allow for the encoding of items located between lines of the original text flow, the offset of the item is given with a resolution of 0.5 lines, positive numbers indicating an upward offset and a negative numbers a downward one.<sup>54</sup> In the case of the text block and the left and right margins, the origin of the offset is the point of reference or logical location of the item, i.e. the line on which the element representing the displaced item is inserted. In the case of the top and bottom margins, the origin is the vertical centre point of the margin in question, the maximum offsets thus being the half the height of the margin in lines.<sup>55</sup> The vertical offset of any vertically displaced items is encoded using the @rend attribute and a value constructed in the following fashion:

@rend	(on <i>textual container</i> elements)
'offset(x)'	offset vertically <i>x</i> lines from the point of reference, where <i>x</i> is ±{int} or ±{int}.5

A conceptually special case of offset is constituted by cases where the item has been displaced to the page preceding or following the one on which the point of reference is located. An example of such a case is a rubrication instruction at the very bottom of a page in a case where the rubricator has deemed the space insufficient and added the rubrication to the beginning—or top margin—of the following page. Although this kind of situation could be handled by a @place attribute value of 'overleaf' or 'opposite', this would preclude the specification of the item's actual location on that page (and also be directionally context-dependent in terms of whether the mark is on the next or previous page). In order to allow for the precise and unambiguous encoding of such items, the rend attribute and a value constructed as follows has been used:

<sup>54</sup> In order to detach the offset passage from the textual flow and to indicate its horizontal location, the element containing the vertically offset element or its ancestor element will also need to provide appropriate values for the @place and if applicable, @rendition attribute. In other words, vertical offset can be used only in the context of displaced elements (except for drop capitals, for which see subsection 11.5.1).

<sup>55</sup> The relative height of the top and bottom margins in lines are calculated from the absolute height of the margin and the average line height, encoded in the physical description of the manuscript.

@rend	(on <i>textual container</i> elements)
'offset(xpages)'	offset x pages from the page on which the point of reference occurs, where x is ±{int}

Since the offset in this case is not defined in terms of the spatial coordinates of the page, but in terms of the linear structure of the manuscript as a sequence of pages, positive integers refer to pages following the point of reference and negative integers to pages preceding it. For the vertical offset of the item on that page, the vertical origin again depends on the value of the place element. In the case of the top and bottom margins, it is the vertical centre of that margin as usual, but in the case of the text block and the left or right margin it corresponds to the first line of text on the page.<sup>56</sup>

A further special case of displacement is that of *rotation*, which occurs mainly in the case of notes or additions in the side margins of the page. Items that have been written vertically up or down the page are indicated using a value constructed as follows on the @rend attribute:

@rend	(on <i>textual container</i> elements)
'rotate(x)'	rotated x degrees clockwise from horizontal, <sup>57</sup> where x can have a numeric value that is a multiple of 45 (e.g. '90' (down the page), '180' (upside down) or '270' (up the page), '315' (diagonally upwards))

In terms of annotating the internal layout of the displaced item, items consisting of a single line of text are treated as ‘in-line’ elements which are not considered to begin a new line of text (and thus do not begin with a <lb/> element), while longer or more complex items that span multiple lines are viewed as ‘block’ elements whose layout is indicated normally (i.e. beginning each line with a <lb/>). In the case of notes or additions consisting of several paragraphs, lists or other more complex textual structures, the appropriate elements are used within the displaced element to indicate this structure, along with the required layout annotation.

### 11.4 Genetic features of the text

This heading, which is not intended to imply any specific theoretical position, is here used simply to refer to those aspects of the manuscript that explicitly have to do with its genesis, namely the identification and recording of the differing scribal origins of different textual components, the *emendations*—additions, deletions and alterations—made to the text after its initial writing, and various kinds of annotations added to the document—whether connected to its text or not. While naturally a highly interpretive activity involving editorial judgement, the annota-

<sup>56</sup> The same vertical origin is also used for ‘detached’ notes in the side margins or the text block that are not attached to a specific point in the text but to the entire page (see *Detached notes* in subsection 11.4.3).

<sup>57</sup> The point around which the element is rotated is the position used for aligning it horizontally, determined by the alignment component of its @rendition value (i.e. *l\_al*, *r\_al* or *c\_al*, which is the default value), and vertically in the middle of the displaced item (relevant mostly for items consisting of multiple lines of text or large script). Any offset values have also been calculated to reflect this.

tion describing these phenomena is here seen to belong to the descriptive layer, being ontologically parallel to the graphemic transcription of the textual content itself.

### 11.4.1 Scribal hands

In the context of a linguistic edition—as well as many other types of editions—the documentation of the different scribal hands that have contributed to the physical realisation of the textual content of a document is essential for distinguishing textual elements originating from different time periods and different language users. While the *TEI Guidelines* define the @hand attribute precisely for this purpose, it is unfortunately only allowed on a small subset of textual containers, namely those documenting additions and deletions (<add>, <del>, <addSpan>, <delSpan>, <restore>, <subst>), intentional effacement (<damage>, <damageSpan>, <gap>, <unclear>) or text-critical readings (<lem>, <rdg>, <rdgGroup>). Since the intention of this edition is to document the hand of each individual segment of text in the original manuscripts, use of the @hand attribute—moved to a private namespace as @pd:hand—has here been extended to cover all textual container elements (see subsection 11.2.6), the <del> element and the <hi> element.<sup>58</sup> In the case of all textual container elements, the attribute is considered to define the hand of the textual content, while in the case of <del> and <hi>, it refers to the hand in which the deletion is indicated (e.g. by a strikethrough or subpunction) or the segment highlighted (e.g. by touching it up in red pigment or underlining it).

The scribal hands used in each of the manuscript versions are documented in the <physDesc> (physical description) section of the manuscript description (see above) in the <teiHeader>. The description of each hand is encoded as a single <handNote> element, consisting of a series of descriptive attributes defined in the *TEI Guidelines* (TEI Consortium 2014: 369-71, 748-9) and a prose description indicating the parts of the manuscript where the hand is used and the specific features for which it has been used. The amount of information encoded about the hand using the attributes can vary significantly depending on how much is known of the hand. Some features of the hand—such as the scribe to whom it belongs may simply be unknown—and others—like the script used—may be unclear due to the limited scope of the hand (e.g. a hand in which only numerals have been written), in which case the value 'unknown' is used for the relevant attribute. At the minimum, each <handNote> element is provided at with a unique @xml:id identifier,<sup>59</sup> whose value is referred to by the @hand attributes representing segments of the text written in that hand. In addition to an identifier, most hands are characterized by the @script attribute, which takes as its value a suitable selection of the following terms, separated by whitespace:<sup>60</sup>

<sup>58</sup> A feature request has also been submitted to the TEI for including the @hand attribute for at least the more common of these elements (<<https://sourceforge.net/p/tei/feature-requests/480/>>) in the *TEI Guidelines*.

<sup>59</sup> For hands occurring in multiple parts of a composite manuscript (MSS Ad or Ad), the entries for the hand after the first one do not have an @xml:id but use the @sameAs attribute to refer to the @xml:id indicated on the first instance of the <handNote> element.

<sup>60</sup> The attribute may include terms from both of the categories, the minimum being one of the terms for the script family (including 'unknown').

<b>Family</b>	'textura', 'anglicana', 'secretary', 'anglicana-secretary', 'secretary-anglicana', 'italic', 'mixed', 'modern', 'unknown'
<b>Grade</b>	'formata', 'media', 'cursiva', 'hybrida',

Since the identity of any the scribes responsible for the six *PD* versions edited here is not known to any degree of certainty, the scribes are identified merely by abstract labels (e.g. 'recipe\_scribe') using the @scribe attribute (instead of the @scribeRef attribute referring to defined persons). The @medium attribute is used to characterize “the tint or type of ink, e.g. *brown*, or other writing medium, e.g. *pencil*” (TEI Consortium 2014: 370) for all those hands for which it can be determined. The following values—which are necessarily subjective and relative in nature—are used for describing the medium in the present edition.<sup>61</sup>

<b>@medium</b>	(on the <handNote> element)
'brown'	indicates a clearly medium-brown ink
'light-brown'	indicates a lighter, yellowish-brown ink
'red-brown'	indicates a warm, reddish brown ink
'dark-brown'	indicates a clearly darker, almost black but still clearly brownish ink
'grey-brown'	indicates a very cold brown ink or gray ink with a slight brown hue
'grey'	indicates a medium gray ink, clearly not black but with no brown hue
'light-grey'	indicates a lighter, not quite opaque ink with no brown hue
'dark-grey'	indicates a dark gray, but not quite black ink with no brown hue
'black'	indicates a solid black ink
'purple-black'	indicates a black ink with a purplish or bluish hue (especially at the edges)
'red'	indicates a red pigment used for rubrication
'blue'	indicates a blue pigment used for rubrication
'pencil'	indicates a pencil
'variable'	indicates that the colour and strength of the ink varies inconsistently
'faded'	indicates that the ink has faded significantly, making its original colour difficult to determine

Finally, all hands found in the manuscript are characterized in terms of their @scope, using the value 'sole' for those hands that are the only one used in the manuscript, 'major' for those that are used for most of the manuscript, and 'minor' for all those that are used only occasionally in the manuscript. The <handNote> describing the principal scribal hand of the edited text, described in considerably more detail, is identified by a <dimensions> element occurring as its first child and containing two <dim> elements describing the dimensions of the hand. The first of these has a @type of 'average-minim-height' and encodes the average height, in millimetres, of the minims in the text (the 'x-height' in print terminology) using the @unit and @quantity attributes, while the second has a @type of 'average-20-minim-width' and encodes the average width of 20 minims in the same manner.<sup>62</sup>

<sup>61</sup> These values are also used to visually represent the hands in the diplomatic transcripts included in appendices B and D.

<sup>62</sup> Since different hands vary widely in their dimensions, these measurements are intended to be used for the calibration of the diplomatic visual presentation of the edited text.

The prose description of the hand formulates all of the aforementioned aspects of the hand in running prose and expresses any qualifications or ambiguities that pertain to the hand. In cases where the identification or description of the hand is based on a secondary source, this source is identified by a bibliographic reference. For all those hands which occur in the edited text, the prose description contains a noun phrase, separately annotated as a <seg> of @type 'handname', which identifies the hand in a way as to be used as the 'name' of the hand in automatically generated textual notes. *XML Example 10* contains examples of <handNote> elements describing a principal scribal hand and a later annotator's hand.

---

**XML Example 10:** *Examples of <handNote> elements describing scribal hands in the TEI header.*

---

```
<handNote xml:id="hand_scribal2a" script="secretary-anglicana" medium="dark-brown"
  scribe="recipe_scribe" scope="major">
  <dimensions>
    <dim type="average-minim-height" unit="mm" quantity="1.5"></dim>
    <dim type="average-20-minim-width" unit="mm" quantity="22"></dim>
  </dimensions>
  <p><seg type="handname">The original scribal hand</seg> used for the recipe collection [...]</p>
</handNote>
...
<handNote xml:id="hand_annotator2" script="secretary" medium="black" scribe="annotator2" scope="minor">
  <p><seg type="handname">An early modern secretary hand in black ink</seg> that has [...]</p>
</handNote>
```

---

Within the transcription itself, the principal hand responsible for the majority of the collection is indicated by using the @hand attribute on the <text> element containing the entire collection, as described above. Any parts of the text that are written in a different hand are indicated by a @hand attribute with the appropriate value on the relevant textual container. Thus, the hand responsible for any specific word in the text is indicated by the @hand attribute value of the nearest textual container ancestor.

In addition to textual containers, the @hand attribute is also used in the elements <hi> and <del>. While its use is identical to textual containers, the semantic interpretation is slightly different: on these elements the @hand attribute does not indicate responsibility for the textual content of the element, but rather indicates the hand responsible for the highlighting or the deletion affecting the textual content. If no @hand attribute is present on these elements, the highlighting or deletion is understood to be performed in the same hand that is responsible for the textual content (i.e. that indicated by the nearest textual container ancestor).

## 11.4.2 Emendation

In the present context, *emendation* is used to refer not to the activity of the editor, but rather various changes—additions, deletions, substitutions and reorderings—that have been performed on the text after its initial writing and are somehow indicated in the manuscript.<sup>63</sup> Although the interpretation of the various markings

---

<sup>63</sup> An important thing to note here is that the writing of the document is here understood as a spatio-temporal process: the point of writing moves through the manuscript as the scribe sets pen to paper or parchment, and the moment of initial writing is always later for items down the textual flow, that in the European context conventionally progresses from left to right and top to bottom. This means

found in the manuscript as signifying emendation is naturally an analytical activity, the annotation of emendation is nevertheless considered to belong to the category of descriptive annotation, as it records phenomena that are actually present in the document. In accordance with this definition, mistakes like omissions or repetitions that have not been indicated as such in the document are not annotated even if they seem obvious, since the annotation of emendations is intended to reflect the perceptions and interventions of earlier users of the document, not the editor’s opinions. Emendations made by the original author and later correctors, annotators and readers are treated identically, being distinguished only by the `@hand` attribute, just like any other contributions to the manuscript’s textual content.

Deletions

All of the manuscript texts in this edition contain passages—from single characters to whole lines of text—that have been for one reason or another been considered superfluous or erroneous and marked for omission from the text, either by the original scribe him- or herself or by later correctors or annotators. Passages in the manuscript text that have been “deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, annotator, or corrector” (TEI Consortium 2014: 82) are annotated using the `<del>` element. In order to represent the different ways manuscript content can be deleted, the `@type` attribute is used to classify the deletions into four categories based on the method of deletion:

@type	(on the <code>&lt;del&gt;</code> element)
'marked'	text marked for deletion in some way
'erased'	text removed physically, usually by scraping with a sharp knife
'overwritten'	text overwritten by an addition
'implicit'	text implicitly removed as a result of some other change

Of these four types, the marked variety is on average the most common in the *Potage Dyvers* texts, accounting for roughly half of all deletions and including *expunction*, *underlining*, and *cancellation* using a variety of different kinds of pen strokes.<sup>64</sup> Since medieval scribes and annotators used a variety of different means for marking passages for deletion, all deletions of the `@type` 'marked' use the attribute `@rend` with one (or more)<sup>65</sup> of the following values—defined for this edition—to describe the different visual methods conventionally used for indication deletion:

that an emendation to an ‘earlier’ part of the text could well—and frequently does—pre-date the initial moment of writing of some other, later, part of the text, and the definition of emendation used here does not require that it occur after the *completion* of the entire document.

<sup>64</sup> While the term *cancellation* properly applies only to the deletion of longer passages “by means of criss-cross lines (*cancelli* means ‘lattice’ or ‘trellis’)” drawn across them, “it is also applied to the crossing-out of individual words” (Hector 1966: 50).

<sup>65</sup> In cases where a single act of deletion is performed (in a single hand) through several different markings (e.g. subpunction and strikethrough), the `@rend` attribute has been given all of the relevant values, separated by a space.



@rendition	(on <del> elements of type 'marked')
'#singlestrike'	marked for deletion by a single horizontal strikethrough
'#doublestrike'	marked for deletion by a double horizontal strikethrough
'#verticalstrike'	marked for deletion by a single vertical strikethrough
'#diagstrike'	marked for deletion by a single diagonal strikethrough descending to the left (/)
'#doublediagstrike'	marked for deletion by double diagonal strikethrough descending to the left (//)
'#revdiagstrike'	marked for deletion by a single diagonal strikethrough descending to the right (\textbackslash)
'#underline'	marked for deletion by underlining with a solid line
'#subpunct'	marked for deletion by dots placed below the text line

Text that has been 'erased' has been physically removed from the page (either completely or incompletely), usually by scraping off the top layer of the paper or parchment. As Hector (1966) points out, erasure is “obviously more suited to parchment, with its close texture and stout substance, than to paper” (49–50), as it can withstand the scraping without tearing and can afterwards be polished with a piece of ivory to restore a smooth surface.<sup>66</sup> This usually means that the deleted text is no longer discernible and is annotated using the <gap> element (see subsection 11.6.3), although in some cases the erasure has not been complete and the text is annotated as merely unclear (see subsection 11.6.2).

While erasures and marked deletions are explicitly discernible in the document, the effects of emendations are not necessarily confined to the site of the visible emendation, but can also radically alter the effect and interpretation of passages around them (Shillingsburg 1986: 45). For this purpose, the present edition defines annotations also for 'overwritten' and 'implicit' deletions, which are not in themselves indicated explicitly but occur indirectly as the result of another emendation. Deletions that are the result of *overwriting* do not constitute an independent act but rather occur as the indirect result of an overwriting addition. For this reason they always occur together with an addition, forming a *substitution* (see *Substitutions*). Depending on the circumstances, the overwritten or *obliterated* text can be either totally obscured by the addition, somewhat unclear or totally legible, and even when both the addition and the deleted content are visible, it is often difficult to determine which character is intended to replace which (Hector 1966: 50). An *implicit deletion* refers to a passage, whose removal from the text has not been explicitly marked, but rather implied by some other, explicit, change made to the text. Like overwritten deletions, also implicit deletions always occur together with an addition—either implicit or explicit—as a part of a substitution. Furthermore, if the corresponding addition is an implicit one, the implicit substitution is by definition accompanied by some form of explicitly marked emendation that ‘triggers’ the implicit substitution.

XML Example 11 contains an example of a typical implicit deletion (and addition) triggered by an explicit emendation, or in this case, a series of emendations.

<sup>66</sup> Although, as Hector (1966) observes, even the polished surface often remains more or less absorbent, resulting in ink applied to it spreading slightly.

<sup>67</sup> Tagging related to abbreviation has been removed for the sake of clarity.

---

**XML Example 11: Implicit substitution triggered by explicit emendation.**<sup>67</sup>


---

```

<w>he
  <subst>
    <del type="overwritten" hand="#scribal">n</del>
    <add type="overwritten" hand="#scribal">ro</add>
  </subst>
  <subst>
    <del type="implicit" hand="#scribal">n</del>
    <add type="implicit" hand="#scribal">u</add>
  </subst>
  <add type="inline" hand="#scribal"><am><g ref="#tittle"/></am></add>
  <del type="marked" rend="subpunct" hand="#scribal">e</del>
</w>

```

---

In the example, the word *henne* ('hen') has been emended into the word *heroun* ('heron') by writing the characters *ro* over the letter *n*, adding a superscript *tittle* above the second letter *n* of *henne* and marking the final *e* for deletion by placing a punctus (dot) below it. Taken by themselves, these changes would result in the word *heroñ*, which is unlikely, since the only possible interpretation for the *tittle* added over the *n* would be either *e* (in which case the scribe could just have left the original final *e* in place) or an otiose flourish (in which case he could have just not added it). Given that the glyphs for *n* and *u* both consist of two minims and are in most cases indistinguishable from each other except for the context, a more likely interpretation is that the emended word is actually *heroū* with the *tittle* standing for the nasal *n*, the *n* having been *implicitly* turned into a *u* as the result of the contextual change.

All deletions—even implicit ones—are considered to be the result of human intervention and thus have a *@hand* attribute whose value indicates the hand in which the deletion (or the emendation that triggered the implicit deletion) was made. See subsection 11.4.1 for more information on the annotation of scribal hands.

### Additions

In addition to deleted passages, all of the texts also contain additions made after the initial writing of the text, either by the original scribe (in which case they may have been made even before the whole text was complete) or by a later corrector, annotator or even a reader. Additions are here defined as any text that has been inserted 'into the logical textual stream'<sup>68</sup> after the writer had passed the point of insertion in his initial writing, i.e. written at least one of the characters that follow the point of insertion in the textual stream, before making the addition.<sup>69</sup> Additions to the text are encoded by inserting an *<add>* element at the *logical point of insertion*, i.e. the point in the text flow where the added material logically

<sup>68</sup> This distinguishes it from a *<note>*, which can also be a later addition, but is not inserted into the textual stream but parallel to it.

<sup>69</sup> This definition means that a situation in which the scribe writes makes a mistake in the middle of a word, writing the wrong letter, but immediately notices this, deletes the erroneous letter and writes the correct one to the right of it, and moves on does *not* involve an addition (although it does involve a deletion). However, a situation in which the scribe misspells one letter in the middle of the word, notices this after finishing the word, goes back to delete the offending letter, and adds the correct one on top of or above the erroneous one, *does* constitute an addition in addition to the deletion, and is also annotated as a *substitution* (see below).

belongs, regardless of the actual physical location of the addition. Additions are categorized by using the following values of the @type attribute—defined for this edition—based on their relationship to the point of insertion:

@type	(on the <add> element)
'inline'	addition made in a blank space on the line in the text flow (usually at the end or beginning of line)
'displaced'	addition displaced from the point of insertion (location indicated as with other displaced text)
'overwritten'	addition overwriting earlier, deleted, text (both always enclosed within a <subst> element)
'implicit'	addition implicitly made by some other change

For 'inline', 'overwritten' and 'implicit' additions, the physical location of the addition on the manuscript page is implied (i.e. at the point on the element in the text flow, on the preceding deleted word, or not physically present). For displaced additions, the location of the addition is indicated by the @place, @rendition and @rend attributes just like other displaced elements, as described in subsection 11.3.5. In some cases, the point of insertion for the addition has been explicitly indicated by a symbol, most commonly a *caret*. These kinds of additions have been indicated by a @rend value of 'marked(sym)', where *sym* is either a Unicode character or a pointer to a <char> element in the header (e.g. “'#asterisk'”, see subsection 11.5.3) representing the symbol used to indicate the point of insertion.<sup>70</sup>

### Substitutions

Emendations involving a deletion and a corresponding addition are together seen to constitute a *substitution* and explicitly annotated as such. For an emendation to qualify as a substitution, the deletion and addition involved in it have to be: 1) spatially concurrent, i.e. involving the same point in the text flow, 2) made in the same hand, and 3) intended as mutually exclusive. Substitutions are encoded by placing the <subst> element—defined for this purpose by the *TEI Guidelines*—around the constituent <del> and <add> element.<sup>71</sup> No attributes are used on the <subst> element, the nature of the substitution being indicated by the attributes of the <del> and <add> elements.

### Transpositions

In addition to additions, deletions and substitutions, scribal emendation can also consist of the reorganization of words or passages in the text without the addition or removal of material. Unfortunately, the current recommendation of the *TEI Guidelines* for this purpose—recently introduced as a part of the new guidelines

<sup>70</sup> Any symbol marking the point of insertion is assumed to be located inline with the surrounding text, and any subscript or superscript is reflected by the character itself.

<sup>71</sup> However, the definition of a substitution used in this edition is stricter than the one explicated in the *TEI Guidelines* (TEI Consortium 2014: 365-7). These stricter limitations are, however, implicit in the TEI definition which states that the substitution should represent “a single intervention in the text”. For the same reason, a substitution in this edition always contains a single addition/deletion pair.

for annotating the genetic history of the document—uses annotation structures (e.g. stand-off links and the concept of *metamarks*) that differ quite significantly from the established structures used for annotating additions, deletions and substitutions, and could not be smoothly integrated with them.<sup>72</sup> For this reason, a simpler system based on inline markup and modelled on the established guidelines for the annotation of addition and deletions and implemented using generic elements defined by the *TEI Guidelines* is used for the present edition.

Since the transposition of elements can be seen to be analogous to the substitution of one passage for another, the present edition uses a structure similar to that used for substitutions, but employing the generic `<mod>` element (of `@type` 'transposition') instead of the `<subst>` element and replacing the `<del>` and `<add>` elements with a series of `<seg>` elements of `@type` 'transposed', each representing a transposed segment of text. The actual order of the `<seg>` elements and their contents within the `<mod>` should reflect the original order before the transposition, while the transposed order is indicated by providing each of the `<seg>` elements with an `@n` attribute indicating their emended order. The `@rend` attribute is used on the `<seg>` elements to indicate the markings used by the scribe to indicate the new order—as well as the location of the markings—using the following values, defined for this edition:<sup>73</sup>

@rend	(on a <seg> element within a transposition)
'marked(x)above'	symbol <i>x</i> placed above the segment to indicate its transposition
'marked(x)below'	symbol <i>x</i> placed below the segment to indicate its transposition
'marked(x)before'	symbol <i>x</i> placed before the segment to indicate its transposition
'marked(x)after'	symbol <i>x</i> placed after the segment to indicate its transposition

The hand in which the transposition has been done is indicated normally using the `@hand` attribute on the `<mod>` element. *XML Example 12* shows a hypothetical example of a transposed segment and its annotation, omitting all markup not related to the transposition.

### 11.4.3 Manuscript notes and *forme work*

While Genette (1997) has defined a *note* as “a statement of variable length (one word is enough) connected to a more or less definite segment of text and either placed opposite or keyed to this segment” (319), emphasising its connection to the text, the present edition adopts a more physically oriented view, seeing notes as

<sup>72</sup> The new recommendation is described in the section “Marking up the Writing Process” (TEI Consortium 2014: 377–84), added to version 2.0 released in December 2011 and intended for the genetic description of documents as physical artefacts, and thus based on a slightly different textual ontology than the main part of the *TEI Guidelines*. The recommended method of annotating transpositions is based on the use of the newly introduced `<metamark>` element to annotate the markings used to indicate the transposition, and of the `<transpose>` and `<listTranspose>` elements to indicate transposed segments through pointers placed elsewhere in the document. Since its use would require extensive changes also to the established method of annotating additions and deletions to keep them conceptually and structurally harmonious, and there are only two instances of transposition in the present edition, it has not yet been implemented in these guidelines, but will be incorporated into the future revised and generalised version of these guidelines, along with revisions to the annotation of additions and deletions.

<sup>73</sup> The symbol used by the scribe can be indicated either as a Unicode character or a pointer to a `<char>` element defined in the TEI header (i.e. the name of the character preceded by a hash).

---

**XML Example 12:** *Transposition of words in the manuscript.*


---

c                    a                    b

order confused word

```
<mod type="transposition" hand="#hand_scribal">
  <seg type="transposed" rend="marked(c)above" n="3">order</seg>
  <seg type="transposed" rend="marked(a)above" n="1">confused</seg>
  <seg type="transposed" rend="marked(b)above" n="2">word</seg>
</mod>
```

---

a feature of the *document* rather than the *text*. This means that the concept of manuscript annotation in its widest sense is here seen to refer to all textual and graphical content that is not part of the main textual stream, whether it is connected to it or not. This view of manuscript notes is very similar to that expressed by the *TEI Guidelines*, which define a note as “any additional comment found in a text, marked in some way as being out of the main textual stream”, and point out that they “may be in a different hand or typeface, may be authorial or editorial, and may have been added later” (TEI Consortium 2014: 103).

Thus the definition of a note includes all marginal scribbles not related to the text itself, marginal or interlinear clarifications of words or passages,<sup>74</sup> all reference markers (e.g. linking two recipes together) added to the margins, and all finding aids and labels such as recipe numbers that are not a part of the original conception of the text. A special category of notes, annotated separately from other notes and discussed separately below, is constituted by what the *TEI Guidelines*—by analogy with printed texts—call *forme work*, annotations that are explicitly connected with the physical makeup of the document, including page numbers, signatures, catchwords, etc.

Although notes are thus by definition something outside the text, many of them are “attached to a specific point or span within a text, which we term here its *point of attachment*” (TEI Consortium 2014: 103). In the present edition, all such notes are transcribed within a <note> element inserted at the logical *point of attachment*. Notes that are associated with larger structural units, such as entire recipes or list items—which includes most of the notes in these manuscripts—are placed within that structural element before its other content.<sup>75</sup> Since notes are by definition something outside of the main textual stream, they are by default *displaced items*, and their location is encoded using the <place>, <rendition> and <rend> attributes as described under subsection 11.3.5.<sup>76</sup> An exception to this is

---

<sup>74</sup> The distinction between a marginal or interlinear note clarifying a word or passage (either in terms of its meaning or simply of its orthography) and a marginal or interlinear addition is that an addition literally adds or replaces something in the text, while a note supplements or comments on something that is already in the text and is not replaced or removed by the note. In practice this means that if a word has been deleted in the text or is judged to be missing from the text, a word added in the margin is annotated as an addition (in the former case forming a substitution), but if the word added in the margin merely replicates—in a more intelligible form—a word in the text, it is annotated as a note.

<sup>75</sup> This means that <note> elements that are the first child of their parent textual container should be understood to be associated with the entire parent element.

<sup>76</sup> In cases where the note consists of several disjointed segments of text, the locations of these segments can be indicated separately by enclosing them within <seg> elements inside the <note> element, with the appropriate @place attribute values (and any accompanying @rendition and @rend

constituted by notes which describe some feature—most commonly damage, but occasionally also ruling or decoration—of the whole page and are thus not localised to any specific point. These kinds of notes are provided with a `@place` attribute value indicating the part of the page to which they relate, but do not have a `@rendition` value indicating their specific location.

Since the category of notes by this definition is quite large and varied, it has been divided into several subcategories based on the basic function of the note. While such a classification can be seen as a form of analysis and thus inappropriate for inclusion in the base data file, it is here seen merely as a way of countering the overly generic nature of the `<note>` element and to raise it to roughly the same level of detail as the other elements used to annotate textual structure. The following values of the `@type` attribute—defined for this edition—are used to indicate the type of each note found in the manuscript:<sup>77</sup>

@type	(on the <note> element)
'gloss'	a note glossing a word or a phrase in the text, i.e. providing a synonym or explanation for it
'label'	a note providing an item in the text with an identifying label, often a number or a name
'marker'	a note marking the location of an item in the text without providing a label for it
'pointer'	a note that indicates a connection between its point of attachment and a reference point somewhere else in the text or even outside of it
'detached'	a note that is not attached to any specific point in the text or to any of its component parts, either because it is totally unrelated or because it relates to the manuscript page in general (see below under <i>Detached notes</i> for more information)
'damage'	a note that contains a representation of manuscript damage not affecting text, located either on the page outside its textual content (in the margins or empty spaces of the text block) or between pages (see section 11.6 for more information)

While the item identified or marked by a 'label' or a 'marker' is indicated implicitly, being the closest suitable ancestor,<sup>78</sup> the point of reference for a 'gloss' or a 'pointer' is indicated explicitly using a `@target` attribute. If the annotation has been made in a hand other than that indicated by its context (usually the hand of the original scribe), this is indicated using the `@hand` attribute.

values) provided for the `<seg>` elements. In this case the `<note>` element should not have a `@place` attribute, indicating merely the existence of the note, the displacement of its components being indicated by the `<seg>` elements.

<sup>77</sup> While many of the note types listed can also be seen to have an analogous element (e.g. `<gloss>`, `<label>`, and `<ref>`), they are much more limited than the `<note>` element in terms of their allowed contexts and not always semantically appropriate.

<sup>78</sup> Often the point of reference is the parent element, but in some cases there may be intervening structural elements—such as the `<front>` of a `<text>` representing an individual recipe—that are not considered to limit the scope of the reference, i.e. a marker or a label located in the `<front>` of a recipe `<text>` is interpreted to refer to the entire recipe rather than just the front, just like a heading contained in a `<head>` element within the `<front>` is considered to refer to the entire text.

Detached notes

A special category of notes is made up by annotations made in the manuscript that are not only outside the main textual stream, but do not refer or relate to any part of the text—or to the physical makeup of the document, which would make them part of the *forme work* (see below). These notes—including things like pen trials, marginal scribbles, notes of ownership and library catalogue numbers—are considered to be outside not only the main text flow, but also of the hierarchical XML structure, not considered to ‘belong’ or refer to their parent element but rather to the manuscript page represented by the preceding <pb> element. They are annotated using a <note> element with a @type of ‘detached’ located at the head of the page right after the <pb> element and any *forme work* elements following it.<sup>79</sup> While distinguished as their own category (since they do not actually contain an annotation in the sense of textual content), notes of @type ‘damage’ are also considered to be *detached* for the purposes of indicating their location.

Forme work

The term *forme work*, borrowed from the world of letterpress printing—where the *forme* is the block of wood or frame used to hold the set type for printing—is used in the *TEI Guidelines* to refer to such information “as page numbers, signatures, or catchwords” (TEI Consortium 2014: 391), i.e. information related not to the textual content but to the functioning of the manuscript as a physical object. While this kind of information, repeated from page to page (or quire to quire in the case of catchwords) and usually contained in the top and bottom margins, can be considered to be a specialised form of *detached notes*, it is encoded separately using the specialized textual container element <fw>. The different types of *forme work* information found in the *Potage Dyvers* manuscripts and annotated in this edition are described using the following values of the @type attribute, defined for this edition on the basis of the suggestions made in the *TEI Guidelines*:

@type	(on the <fw> element)
'pageno'	(page number) sequential numbering of pages, i.e. sides of leaves
'folio'	(folio number) sequential numbering of folia, i.e. leaves
'quire'	(quire number) sequential numbering of quires or booklets
'runhead'	(running head) a running head, usually the title of the collection
'sig'	(signature) a quire signature, identifying the quire and the folio within it
'catch'	(catchword) the first word of the following quire inserted on the last page of the quire

In terms of the document structure, the <fw> element (along with notes of the types ‘detached’ and ‘damage’, described above) is a special case, because while it is a *textual container*, it is also by definition not connected to the logical structure of the text but rather to the physical manuscript page. This means that it should be considered to be outside the normal nesting structure of the XML document and

<sup>79</sup> The physical location of these notes, which are by definition *displaced items*, is indicated using the normal procedure described under subsection 11.3.5. The vertical location of detached notes located in the text block or the side margins is indicated in relation to the first line on the page, similarly to items that have been displaced over a page division.

thus not associated with its parent element but rather with the preceding `<pb>` element.<sup>80</sup> In order to make this association more explicit, all `<fw>` elements are in this edition placed immediately after the `<pb>` element of the page on which they occur. In being outside the main textual flow, *forme work* are by definition displaced items and thus always have the `@place` attribute. Since they are not associated with any point in the text, the offset of `<fw>` elements that have a `@place` value of 'text-block', 'margin-left' or 'margin-right' is expressed relative to the first line of the page.

## 11.5 Visual characteristics

Although the *Potage Dyvers* manuscripts are typical representatives of medieval utilitarian texts in that they are quite plain in appearance, there are some rather modest examples of decoration, often used to highlight significant words in the text or to indicate document structure. Although the *TEI Guidelines* define several different elements for annotating words emphasized or highlighted for different reasons (TEI Consortium 2014: 66–78), the present edition exclusively uses the semantically neutral `<hi>` element to annotate characters, words or phrases that are “graphically distinct from the surrounding text” (TEI Consortium 2014: 67). While the guidelines also allow the annotation of visual features directly on any element using the global `@rendition` and `@rend` elements, the present edition *always* uses the `<hi>` element, even in cases where it coincides exactly with a text container element in order to separate the descriptive annotation of visual features from the analytical annotation of textual structure.<sup>81</sup>

### 11.5.1 Highlighting, decoration and emphasis

The visual characteristics of highlighting, decoration and emphasis are encoded using the `@rendition` and `@rend` attributes on the `<hi>` element.<sup>82</sup> The use of both the `@rendition` and the `@rend` attributes reflects the fact that decoration and emphasis in manuscript texts can take various forms, and not all of them can be treated identically from an encoding point of view. Since `@rendition` takes as its value a pointer to a visual rendition that needs to be defined in the `<encodingDesc>` part of the TEI header (see subsection 11.1.3), it is not ideally suited to those forms of highlighting that involve some kind of qualitative or quantitative variables. Forms of highlighting that can involve additional variables such as size, colour or type, are therefore encoded using the `@rend` attribute which can take any number of whitespace-separated tokens as its value, avoiding the need to explicitly define a rendition for all possible values of a variable. The following

<sup>80</sup> One possible solution for this structural anomaly would be to make the `<pb>` element non-empty and place all of the *forme work* (and detached note) elements inside it. This, however, would require deviating rather drastically from the content model of the *TEI Guidelines*.

<sup>81</sup> In these cases, the `<hi>` element is nested directly inside that text container element, becoming its only child element.

<sup>82</sup> Although these attributes are used also on other elements for other purposes, the `<hi>` element is the only one on which they should be understood to indicate highlighting of any kind. The other uses of the `@rendition` and `@rend` attributes are discussed under subsections 11.3.5 and 11.4.2.



values of these attributes—defined for this edition—are used on the <hi> element to indicate visual highlighting of text:

<b>@rendition</b>	(on the <hi> element)
'#sub'	characters that have been written as subscript but do not signify an abbreviation (see subsection 11.5.3 below)
'#sup'	characters that have been written as superscript but do not signify an abbreviation (see subsection 11.5.3 below) <sup>83</sup>
'#sc'	(small capitals) majuscule letters that equal minuscules in height
'#lm'	(large minuscules) minuscule letters that equal majuscules in height
'#da'	(decorative ascender) characters that have a decorative ascender of unusual length extending above them
'#dd'	(decorative descender) characters that have a decorative descender of unusual length extending below them
'#red'	characters that have been written in red pigment
'#blue'	characters that have been written in blue pigment

<b>@rend</b>	(on the <hi> element)
'hl(col)'	(highlighting) characters that have been highlighted by a stroke of coloured pigment, where <i>col</i> indicates the colour used ('red' or 'blue' in the present edition)
'ul' or 'ul(col)'	(underlining) text that has been highlighted by underlining it; if the underlining is in coloured pigment, <i>col</i> indicates the colour used ('red' or 'blue' in the present edition)
'fx' or 'fx(col)'	(framing) span of text text that has been highlighted by framing it with strokes on one or more of its sides, <i>x</i> indicating the side ( <i>t</i> for top, <i>b</i> for bottom, <i>l</i> for left and <i>r</i> for right) <sup>84</sup> ; if the framing is in coloured pigment, <i>col</i> indicates the colour used ('red' or 'blue' in the present edition)
'dc(#)'	(drop capital) a drop capital # lines in height, its top aligned with the headline of the line it is located on <sup>85</sup>
'lc(#)'	(large capital) a large capital # lines in height, standing on the line
'size(#)'	text that has been written with characters # times the size of the 'default' script size (to the nearest 0.25 in sizes under 1 and to the nearest 0.5 in sizes over 1)
'script(family)'	text that has been written using a different script family from the one normally used by the hand for the purposes of decoration or emphasis, where <i>family</i> indicates the script family used (only <i>uncial</i> in the present edition) <sup>86</sup>

<sup>83</sup> The two uses of superscript characters have been annotated separately because, first of all, the *TEI Guidelines* do not allow for the use of the <hi> element within the <am> (abbreviation marker) element (see section 11.7), and secondly, because several superscript abbreviation symbols are different enough from their inline counterparts—both graphetically and functionally—that they need to be seen as independent graphemes (see subsection 11.5.3 below).

<sup>84</sup> A distinct value, separated by whitespace, is used for each side of the text, i.e. the value for a span of text framed on all sides would be 'ft fl fr fb' (order of the individual values is not significant).

<sup>85</sup> In cases where the top of the drop capital is not aligned with the top of the containing line, an 'offset(x)' value on the same @rend attribute can be used to indicate the vertical offset.

<sup>86</sup> This keyword refers to the rounded capitals “of something like uncial form” developed in the 12<sup>th</sup> century, sometimes also called ‘Lombardic’.

To allow for the annotation of cases where the hand responsible for the highlighting is different from the hand responsible for the text, this edition extends the *TEI Guidelines* by providing the `<hi>` element with the attribute `@hand`. It should be noted that the `<hi>` element is not a textual container and thus does not serve as the parent of the contained text in terms of attribute inheritance, the `@hand` attribute referring only to the hand in which the highlighting was added.<sup>87</sup> In cases where the highlighted segment is also written in a different hand from its surrounding text, the `<hi>` element is surrounded by a `<seg>` element with the appropriate value of the `@hand` attribute.

### 11.5.2 Images and ornaments

While none of the *Potage Dyvers* manuscripts contain any illustrations, they do exhibit a number of other graphical and ornamental features. Ornaments are here defined as graphical features which do not have textual or iconic function but are present only for decorative purposes. Examples of ornaments include line-fillers and ornamental horizontal bars or vertical columns, which are intended merely to fill up blank spaces on the page. A defining characteristic of an ornamental feature is their redundancy in terms of textual function: simple empty space could be used to fulfil the same text-organizing function. All graphical elements, both illustrative figures and ornaments are encoded into the transcription using the `<graphic>` element defined by the *TEI Guidelines* for encoding inline graphics (TEI Consortium 2014: 109-11).

The visual appearance of these graphical elements is described in the header using SVG, an XML language for representing two-dimensional graphics (Dahlström et al. 2011). This is accomplished by inserting an `<svg:svg>` element defining a rough representation of the shape of each ornament into a single `<decoNote>` element of `@type 'svg'` within the `<decoDesc>` element of the manuscript description (see subsection 11.1.6). The `<svg:path>` elements that make up the shape should describe only the shape of the ornament, its size and color being defined by the `@rendition` and `@rend` attributes of the ancestor elements of each `<graphic>` element. Each `<svg:svg>` element has an `@xml:id` attribute value uniquely identifying it within the document, which is referred to by the `@url` attribute of the `<graphic>` element. An approximation of the size of the ornament in relation to the surrounding text is encoded using the `@width` and `@height` attributes, expressed using the CSS length unit *em*.<sup>88</sup> If the ornament is drawn in a different hand from that of the surrounding text, the `@hand` attribute is used to indicate this.

<sup>87</sup> In cases where the highlighting is identical with the text itself (including by definition all cases except for underlining and highlighting stroke), the `hand` attribute should not be used on the `<hi>` element but on the enclosing textual container element, from which also the `<hi>` element inherits it.

<sup>88</sup> The *em* is a relative unit of length and is defined as the vertical space needed for the tallest letter of the current font. For vertical measurements, one *em* corresponds to one line (minimum height corresponding to 0.5em), and for horizontal measurements, one *em* corresponds roughly to two characters.

### 11.5.3 Special symbols

Medieval manuscripts contain a variety of graphemes in addition to those strictly considered ‘letters’ or alphabetic characters. Although the Unicode standard (The Unicode Consortium 2011) defines graphemic encodings for many of these special symbols, the coverage is not yet complete. Many of the special symbols required for the diplomatic transcription of medieval manuscripts that are not covered by the Unicode standard have been defined by the Medieval Unicode Font Initiative (MUFI) workgroup and included in specialist fonts making use of the *Private Use Area* (PUA) reserved for the purpose by Unicode. But since not even MUFI covers all of the required symbols, and the code points used are not yet part of the Unicode standard and are thus liable to change, the use of *Numeric Character References* (NCRs) for them would pose a risk for the long-term sustainability of the edition. Additionally, even some of the characters represented in Unicode characters have multiple alternative representations, which means that even a direct Unicode representation of special characters would thus require detailed documentation and additional precautions to ensure consistency in the transcription (Wittern 2006: 295).

To avoid these problem in the case of the undefined symbols and to maintain internal consistency in encoding, all special characters that are not considered to be alphabetical (see subsection 10.2.2) are represented using the `<g>` (glyph) element defined for this purpose in the *TEI Guidelines* (TEI Consortium 2014: 182).<sup>89</sup> Another approach, adopted by many projects, including the *MENOTA* and recommended for example by Haugen (2004: 2004) and Deegan (2006), would be to use ‘entity references’ (such as `&et;`) to represent special characters. However, this approach has multiple drawbacks: first of all, the definition of entity references is no longer supported by modern schema languages, being a relic of the DTD format, and unlike the `<g>` element, they do not allow for the documentation of the visual appearance and properties of the characters. The approach adopted here also has the benefit of being more flexible in terms of processing, since it is much more convenient and efficient in XSLT to convert elements into characters than characters to other characters in the presentation stage.

The different abbreviation markers, punctuation marks and other special symbols used in the edited manuscripts are listed below. They are also enumerated and formally described in the `<encodingDesc>` part of the TEI header of each base data file, using the `<charDecl>` element defined by the TEI for this purpose (TEI Consortium 2014: 182-91). Each of the special characters is described by a `<char>` element, which provides an `@xml:id`, a name, a description, and a suggested Unicode or MUFI representation for the special character. In the transcription itself, these special characters are encoded using an empty `<g>` element with a `@ref` attribute pointing to the `@xml:id` of the relevant `<char>` element. The present edition defines the following special characters:

<sup>89</sup> Deegan (2006: 367) has recommended an even more conservative approach of restricting character encoding to the ASCII standard. However, the present edition considers the encodings of the *thorn* and *yogh* characters, which are part of the well-established core of Unicode, to be sufficiently stable even though they are not included in the largely obsolete basic ASCII character set.

@xml:id	(on a <char> element)
'et'	(‘Tironian <i>et</i> ’) the brevigraph for Latin <i>et</i> , here used mostly for Middle English <i>and</i> ; <sup>90</sup> Unicode representation used in this edition: ⟨τ⟩ (204A, TIRONIAN SIGN ET);
'tittle'	(‘superscript <i>tittle</i> ’) the most common abbreviation marker in both vernacular and Latin manuscripts (No. 1 in Denholm-Young 1954: 67), consisting of a straight or curved stroke above the preceding letter, used both as a general marker and to indicate the omission of a nasal, or word-finally also ⟨e⟩; Unicode representations used in this edition (depending on the width of the preceding character): <sup>91</sup> ⟨ˆ⟩ (F00B, COMBINING MEDIUM-HIGH MACRON WITH FIXED HEIGHT (PART-WIDTH)), ⟨˘⟩ (F00D, COMBINING MEDIUM-HIGH MACRON WITH FIXED HEIGHT (FULL-WIDTH));
'sub-3'	(‘subscript 3 <i>et</i> ’) a brevigraph shaped like a number 3 lowered below the baseline, here used exclusively in Latin passages for <i>ue</i> (following the letter <i>q</i> ), <i>et</i> or <i>us</i> ; Unicode representation used in this edition: ⟨₃⟩ (A76B, LATIN SMALL LETTER ET);
'tall-9'	(‘ <i>con</i> ’) a brevigraph shaped like a tall number 9, here used for <i>us</i> (in both Latin and Middle English) and <i>es</i> (in Middle English); Unicode representation used in this edition: ⟨9⟩ (A76F, LATIN SMALL LETTER CON);
'tailstroke'	a diagonal or vertical hairline stroke descending from the tip of the rightmost stroke of a word-final letter, which can either be an otiose flourish or signify the omission of a word-final ⟨e⟩; <sup>92</sup> Unicode representation used in this edition: the replacement of the preceding letter (c, d, f, g, k, l, m, n, r, s or t) with a tail-stroked version defined by the MUFI or by the Unicode standard <sup>93</sup> ;

<sup>90</sup> Although this symbol is often equated with and represented by an ampersand, the printer’s ampersand is in fact an adaptation of an ancient *et* ligature.

<sup>91</sup> When occurring with specific characters, namely ⟨i⟩, ⟨j⟩, ⟨m⟩, ⟨l⟩ or ⟨ll⟩ they are represented by a special precomposed character—which also replaces the character with which the tittle occurs—defined by the MUFI or by the Unicode standard, namely ⟨i̇⟩ (012B, LATIN SMALL LETTER I WITH MACRON (Unicode)), ⟨j̇⟩ (E554, LATIN SMALL LETTER J WITH MEDIUM-HIGH MACRON (ABOVE CHARACTER)), ⟨ṁ⟩ (E5D2, LATIN SMALL LETTER M WITH MEDIUM-HIGH OVERLINE (ABOVE CHARACTER)), ⟨l̇⟩ (A749, LATIN SMALL LETTER L WITH HIGH STROKE (Unicode)), or ⟨ll̇⟩ (1EFB, LATIN SMALL LETTER MIDDLE-WELSH LL (Unicode)). These precomposed characters are used simply to avoid the problems of positioning involved using the combining characters with unusually wide or narrow characters.

<sup>92</sup> Denholm-Young (1954: 68) equates this symbol with the ‘tailed loop’ and lists it as No. 10 in his classification of abbreviation markers, but it is here treated separately in accordance with the principles outlined in subsection 10.2.4.

<sup>93</sup> I.e. ⟨c̣⟩ (F198, LATIN SMALL LETTER C WITH A CURL), ⟨ḍ⟩ (F193, LATIN SMALL LETTER D WITH A CURL), ⟨f̣⟩ (F194, LATIN SMALL LETTER F WITH A CURL), ⟨g̣⟩ (F196, LATIN SMALL LETTER G WITH A CURL), ⟨ḥ⟩ (F195, LATIN SMALL LETTER K WITH A CURL), ⟨ḷ⟩ (A772, LATIN SMALL LETTER LUM (Unicode)), ⟨ṃ⟩ (A773, LATIN SMALL LETTER MUM (Unicode)), ⟨ṇ⟩ (A774, LATIN SMALL LETTER NUM (Unicode)), ⟨ṛ⟩ (A775, LATIN SMALL LETTER RUM (Unicode)), ⟨ṣ⟩ (F1A8, LATIN ABBREVIATION SIGN SCRIPT S WITH STROKE), or ⟨ṭ⟩ (F199, LATIN SMALL LETTER T WITH A CURL). For tail strokes accompanying other characters—⟨e⟩ and the Tironian *et* ⟨τ⟩, an ad-hoc representation by a subscripted slash (/) is currently the only option.

<b>'loop'</b>	a looped stroke curving upwards and around, departing from the rightmost stroke of a letter, usually at or near the headline (Sign No. 9 in Hector 1966), used here mostly to signify the omission of word-final <i>-e</i> , <i>-er</i> or <i>-es</i> , but also as a general sign of abbreviation; <sup>94</sup> no Unicode or MUFI representation—represented in this edition by ⟨ <sup>°</sup> ⟩, a custom symbol at Unicode PUA code point F15F;
<b>'tailed-loop'</b>	a looped stroke similar to the one described by the value 'loop' but crossing itself and extending downwards in a vertical tail (Sign No. 9 in Hector 1966 and No. 10 in Denholm-Young 1954), used word-finally to signify the omission of <i>-es</i> , <i>-is</i> , <i>-ys</i> or “even plain <i>-s</i> ” (Hector 1966: 39); Unicode representation: ⟨ <sup>†</sup> ⟩ (A76D, LATIN SMALL LETTER IS);
<b>'hook'</b>	a superscript stroke in the shape of a hook or a reverse question mark (No. 2 in Denholm-Young 1954: 67), located directly above the preceding letter, here used mostly to indicate the omission of <i>er</i> either word-medially or word-finally; Unicode representation used in this edition: ⟨ <sup>ς</sup> ⟩ (035B, COMBINING ZIGZAG ABOVE);
<b>'tailed-hook'</b>	a version of the hooked stroke (also No. 2 in Denholm-Young 1954: 67), where the hook is connected to the rightmost strike of the letter below by a long, curved stroke, which is actually not a tail but rather an approach stroke; no Unicode or MUFI representation—represented in this edition by ⟨ <sup>⸏</sup> ⟩, a custom symbol at Unicode PUA code point F14E;
<b>'sup-2'</b>	(‘superscript round <i>r</i> ’) a superscripted rotunda <i>r</i> , resembling a small Arabic numeral 2, used here mostly to signal the omission of <i>ur</i> ; no Unicode representation—MUFI representation used in this edition: ⟨ <sup>ˆ</sup> ⟩ (F153, COMBINING ABBREVIATION MARK SUPERSCRIPT UR ROUND R FORM);
<b>'sup-squig'</b>	(‘superscript squiggle’) a superscript brevisgraph that varies significantly in shape but most commonly resembles a curved stroke with a curl or a loop at the left end, used here mostly to signal the omission of <i>ur</i> ; Unicode representation: ⟨ <sup>˘</sup> ⟩ (1DD1, COMBINING UR ABOVE);
<b>'sup-a'</b>	(‘superscript <i>a</i> ’) a superscript flat-topped <i>a</i> , resembling a letter <i>u</i> with a horizontal stroke across its top (sometimes this top stroke is omitted, resulting in an open-topped variant which is not distinguished as a separate symbol in the present edition), used here mostly as an abbreviation marker to signal the omission of <i>ra</i> or <i>ar</i> , or as a marginal marker; no Unicode representation—MUFI representation used in this edition: ⟨ <sup>ˆ</sup> ⟩ (F1C1, COMBINING ABBREVIATION MARK SUPERSCRIPT RA OPEN A FORM WITH BAR ABOVE);
<b>'sup-d'</b>	(‘superscript <i>d</i> ’) a superscript letter <i>d</i> , used here in contractions of Latin words ending in ⟨ <i>d</i> ⟩ (eg. <i>apud</i> ); no Unicode or MUFI representation—represented in this edition by ⟨ <sup>d</sup> ⟩, the letter ⟨ <i>d</i> ⟩ with superscript formatting;

<sup>94</sup> The use of the loop “as an entirely general mark of abbreviation”, especially in suspensions, is also reported by Hector (1966: 31).

'sup-h'	('superscript <i>h</i> ') a superscript letter <i>h</i> , used here in abbreviations; no Unicode or MUFI representation—represented in this edition by ⟨ <sup>h</sup> ⟩, the letter ⟨h⟩ with superscript formatting;
'sup-i'	('superscript <i>i</i> ') a dotless superscript letter <i>i</i> or a single minim, used here mostly as an abbreviation marker to signal the omission of <i>ri</i> or <i>ir</i> ; no Unicode representation—MUFI representation used in this edition: ⟨ <sup>˙</sup> ⟩ (F02F, COMBINING LATIN SMALL LETTER DOTLESS I);
'sup-m'	('superscript <i>m</i> ') a superscript letter <i>m</i> , used here to indicate the contraction of Latin words ending in ⟨m⟩ (eg. <i>Cap<sup>m</sup></i> for 'capitulum'); no Unicode or MUFI representation—represented in this edition by ⟨ <sup>m</sup> ⟩, the letter ⟨m⟩ with superscript formatting;
'sup-n'	('superscript <i>n</i> ') a superscript letter <i>n</i> , used here for a single Latin abbreviation of an unidentified word; no Unicode or MUFI representation—represented in this edition by ⟨ <sup>n</sup> ⟩, the letter ⟨n⟩ with superscript formatting;
'sup-o'	('superscript <i>o</i> ') a superscript letter <i>o</i> , used here to indicate the contraction of Latin words ending in ⟨o⟩ (eg. <i>A<sup>o</sup></i> for 'anno'); no Unicode or MUFI representation—represented in this edition by ⟨ <sup>o</sup> ⟩, the letter ⟨o⟩ with superscript formatting;
'sup-t'	('superscript <i>t</i> ') a superscript letter <i>t</i> , used most frequently to abbreviate the word 'with' ( <i>w<sup>t</sup></i> ) and various words ending in ⟨t⟩, like 'that' ( <i>t<sup>t</sup></i> ); no Unicode or MUFI representation—represented in this edition by ⟨ <sup>t</sup> ⟩, the letter ⟨t⟩ with superscript formatting;
'crossed-p'	('p with crossed descender') a minuscule letter <i>p</i> with a horizontal stroke across the descender, here mostly used to stand for <i>per</i> ; Unicode representation: ⟨p̅⟩ (A751, LATIN SMALL LETTER P WITH STROKE THROUGH DESCENDER);
'crossed-P'	('capital <i>P</i> with crossed stem') a majuscule letter <i>P</i> with a horizontal stroke across the stem, here mostly used to stand for <i>Per</i> ; Unicode representation: ⟨P̅⟩ (A750, LATIN CAPITAL LETTER P WITH STROKE THROUGH DESCENDER);
'looped-p'	('p with loop') a minuscule letter <i>p</i> with a looped stroke extending from the bottom of the lobe to the left of the descender and curling downwards; no Unicode representation—MUFI representation used in this edition: ⟨p̃⟩ (E67D, LATIN SMALL LETTER P WITH FLOURISH)
'looped-s'	('s with looped stroke across stem') a long <i>s</i> with a looped stroke curving down to the right from the middle of the stem, crossing the stem and terminating in a rightward curl; no Unicode representation—MUFI representation used in this edition: ⟨ſ̃⟩ (E8B7, LATIN SMALL LETTER LONG S WITH FLOURISH);
'hyphen'	a symbol that in the late Middle Ages consisted of two slanted hairlines, sometimes used at line end to indicate that the word continues past the line boundary; Unicode representation: ⟨=⟩ (2E17, DOUBLE OBLIQUE HYPHEN);
'dash'	a short horizontal stroke the length of approximately half a character or a single minim, used as a punctuation mark and an annotation symbol; Unicode representation: ⟨-⟩ (2010, HYPHEN);

'mdash'	('m-width dash') a longer horizontal stroke the length of approximately one and a half characters or three minims, used as a punctuation mark; Unicode representation: ⟨—⟩ (2014, EM DASH);
'vline'	('vertical line') a vertical stroke of varying height (usually extends below the baseline and above the headline), used mainly to separate words instead of an inter-word space; Unicode representation: ⟨ ⟩ (007C, VERTICAL LINE);
'vwline'	('vertical wavy line') a vertical wavy line extending roughly from the baseline to the headline, used as a marker symbol with unknown meaning; Unicode representation: ⟨⟩ (2307, WAVY LINE);
'virg'	('virgula') an oblique stroke slanted to the right, a punctuation mark used from the middle of the 15 <sup>th</sup> to the middle of the 16 <sup>th</sup> century for marking a short pause and the precursor of the <i>punctus suspensivus</i> and the modern comma (Hector 1966: 47); Unicode representation: ⟨/⟩ (002F, SOLIDUS);
'virgsusp'	('virgula suspensiva') a combination of a virgule and a medial <i>punctus</i> , used as a punctuation mark; no Unicode representation—MUFI representation used in this edition: ⟨ʃ⟩ (F1F4, VIRGULA SUSPENSIVA);
'wline'	('wavy line') a short wavy line resembling the modern tilde—can occur either alone or in groups forming a longer, continuous undulating line, which has here been annotated as a series of as many wavy lines as there are pairs of rising and falling sections; no Unicode representation—MUFI representation used in this edition: ⟨˘⟩ (F1F9, WAVY LINE);
'punct'	('punctus') a simple point, usually located either on the baseline or the middle of the line, used as a punctuation mark <sup>95</sup> or to separate numbers from the surrounding text; Unicode representation: ⟨.⟩ (002E, FULL STOP);
'colon'	a symbol consisting of two points, one above the other, used as a punctuation mark, traditionally indicating a major medial pause; Unicode representation: ⟨:⟩ (003A, COLON);
'punctelev'	('punctus elevatus') one of the <i>positurae</i> punctuation marks, a point surmounted by a diagonal stroke or an inverted comma shape, traditionally signifying a major medial pause; <sup>96</sup> no Unicode representation—MUFI representation used in this edition: ⟨˙⟩ (F161, PUNCTUS ELEVATUS);
'punctflex'	('punctus flexus') one of the <i>positurae</i> punctuation marks, a point surmounted by a 7-shaped stroke, traditionally signifying a minor medial pause; no Unicode representation—MUFI representation used in this edition: ⟨˙˘⟩ (F1F5, PUNCTUS FLEXUS);
'composit'	('comma positura') a 7-shaped comma-like symbol, originally a critical sign and subsequently a punctuation mark; no Unicode representation—MUFI representation used in this edition: ⟨˙˘˘⟩ (F1E2, COMMA POSITURA);

<sup>95</sup> Since the system of punctuation by *distinctiones* was no longer in use in late Middle English, all points have been annotated the same regardless of their height.

<sup>96</sup> Hector (1966: 46) remarks that in English archival texts the *punctus elevatus* has an indeterminate value, being something “less than that of the modern full stop”.

- 'punctposit'** ('*punctus* with *comma positura*') a compound punctuation mark consisting of a *punctus* followed by a *comma positura*; no Unicode representation—MUFI representation used in this edition: ⟨.,⟩ (F1E4, PUNCTUS WITH COMMA POSITURA);
- 'colmidcomposit'** ('colon with *comma positura*') a compound punctuation mark consisting of a colon followed by a *comma positura* raised to the middle of the line; no Unicode representation—MUFI representation used in this edition: ⟨:·⟩ (F1E5, COLON WITH MIDDLE COMMA POSITURA);
- 'bidotscomposit'** ('two points above a *comma positura*') a compound punctuation mark consisting of a *comma positura* surmounted by a horizontal pair of points; no Unicode representation—MUFI representation used in this edition: ⟨;·⟩ (F1F2, TWO DOTS OVER COMMA POSITURA);
- 'tridotscomposit'** ('three points above a *comma positura*') a compound punctuation mark consisting of a *comma positura* surmounted by three points in an upright triangle formation; no Unicode representation—MUFI representation used in this edition: ⟨;·;⟩ (F1E6, THREE DOTS WITH COMMA POSITURA);
- 'asterisk'** a star-shaped symbol used as a marginal *nota*; Unicode representation: ⟨\*⟩ (2217, ASTERISK OPERATOR);
- 'uparrow'** an upwards-pointing stemless arrow shape, here used as a marginal *nota* (on one occasion); Unicode representation: ⟨^⟩ (2303, UP ARROWHEAD);
- 'subvirg'** ('subscript *virgula*') a short virgule located at the baseline, used to indicate the positions of insertions and the presence of transpositions, similarly to a subscript caret; no Unicode or MUFI representation—represented in this edition by a subscript virgule (,);
- 'doublesubvirg'** ('double subscript *virgula*') a symbol consisting of two short slanted strokes located at the baseline, used to indicate the positions of insertions and the presence of transpositions, similarly to a subscript caret; no Unicode or MUFI representation—represented in this edition by two consecutive subscript virgules (,);
- 'subvline'** ('subscript vertical line') a vertical line located at the baseline, used to indicate the positions of insertions and the presence of transpositions, similarly to a subscript caret; no Unicode or MUFI representation—represented in this edition by a subscript pipe character (|);
- 'doublesubvline'** ('double subscript vertical line') a symbol consisting of two vertical lines located at the baseline, used to indicate the positions of insertions and the presence of transpositions, similarly to a subscript caret; no Unicode or MUFI representation—represented in this edition by two subscript pipe characters (||);
- 'paraph'** a symbol derived from a capital *C* with two vertical strokes (standing for capitulum), here used to indicate the beginning of a new recipe or title; Unicode representation: ⟨¶⟩ (00B6, PILCROW SIGN);



'cross'	a cross symbol consisting of two perpendicular lines of equal length, one horizontal and one vertical, used as a marginal <i>nota</i> ; Unicode representation: ⟨+⟩ (002B, PLUS SIGN);
'xcross'	(‘diagonal cross’) a cross symbol consisting of two perpendicular diagonal lines of equal length, used as a marginal <i>nota</i> ; Unicode representation: ⟨×⟩ (00D7, MULTIPLICATION SIGN);
'refmark'	(‘reference mark’) a diagonal cross with points located between the arms, here used as a <i>nota</i> by a later annotator; Unicode representation: ⟨⌘⟩ (203B, REFERENCE MARK);
'sol'	(‘symbol of sun’) a circled dot, the astrological symbol of the sun, used here as a <i>nota</i> ; Unicode representation: ⟨☉⟩ (2609, SUN).

---

The decision whether to treat a letterform with extra strokes as a letter followed by a separate combining abbreviation marker instead of a special brevigraph is based on the combinatory potential of the said extra strokes. If they can occur on several different letterforms with more or less the same meaning, they have been considered as independent abbreviation markers in themselves and annotated as such following the parent letterform (e.g. the tailstroke). If, on the other hand, the strokes are specific to a single letter or have a different interpretation in connection with different letters, each combination has been annotated as a single brevigraph (e.g. the ‘crossed *p*’).

#### 11.5.4 Differentiating special symbols and graphical elements

In the manuscripts, there are some graphical elements that could—based on their shape—be plausibly argued to be either special symbols or merely graphical markings, including single strokes of the pen—horizontal, vertical or diagonal—and other simple figures. Since there is often great uncertainty as to the status of such a marking in itself, the decision between a special symbol or a graphical element has in these cases been based largely on context.

If a horizontal, vertical or diagonal stroke occurs between two words of the text or at the end of a textual segment, it has been interpreted as a special symbol used as a punctuation mark—a dash, a vertical line or a *virgule*—and annotated accordingly (a <g> element within a <pc> element), but if it occurs in a sequence of similar characters that are merely used to fill up empty space or to form a decorative bar, it is interpreted and annotated as a decorative graphic element. In addition to these rather clear-cut cases, there is also a third, intermediate use of various symbols or markings, namely as notes marking a recipe or some other feature of the text. In this use, described under subsection 11.4.3, the different kinds of crosses, dashes and strokes are interpreted as special symbols, because they serve in this function alongside various symbols—like regular or superscript letters.

While this solution does add a component of editorial interpretation to the low-level transcription of ‘the markings on the page’, it is necessary to draw the line somewhere, as not all graphical markings on the page can be treated as special symbols, and treating them all merely as graphical elements would largely

defeat the point of transcription and render the edition into a rough form of digital facsimile. It should also be noted that the contextual differentiation between graphical elements and special symbols only applies to a limited subset of markings, which are furthermore often graphically distinct even in the original, the decorative form usually being more solid in appearance and often drawn in colour.

## 11.6 Physical condition and legibility

In addition to interventions by correctors, annotators and readers—aiming to improve upon and elucidate the original texts—the original texts as set down by the scribes have also been modified by less benevolent agents. Various kinds of damage, from moisture and holes made by bookworms to partial mutilation or total excision of leaves, have rendered parts of the text unclear to varying degrees or even entirely illegible. This section will describe the annotation of various kinds of damage to the manuscripts and of passages that are either not clearly legible or entirely missing, either due to damage or for some other reason. Since passages of the manuscript text can be unclear or illegible for a number of reasons besides physical damage, this edition adopts the more verbose method of annotation described in the *TEI Guidelines* (TEI Consortium 2014: 375-6), which annotates the damage and its effect to the legibility of the text separately from any damage that may have caused it using `<unclear>` and `<gap>` elements, allowing also the annotation of unclear passages not resulting from damage.<sup>97</sup>

### 11.6.1 Damaged passages

Since physical damage suffered by a manuscript not only affects the legibility of its textual contents but can also provide information about the history of the manuscript and the conditions in which it has been stored, this edition annotates all identifiable instances of physical damage to the manuscript, not only ones which affect the legibility of the text. There are, however, some qualifications to this statement. Since the aim is 1) to document the existence of certain types of damage, and 2) to document the extent to which they affect the text, the edition does not try to annotate the extent of damage outside the text, merely its existence and nature. Thus, for example, water damage extending across the outer margin and into the text block is explicitly annotated only for the text block, the damage to the empty margin being left implicit. However, holes and other highly specific local instances of damage that are limited only to the periphery of the page are annotated even if they do not obscure any text, since they would otherwise escape entirely undocumented.<sup>98</sup>

Physical damage is annotated using the `<damage>` element, as described in the *TEI Guidelines* (TEI Consortium 2014: 373-5). In annotating a damaged stretch of

<sup>97</sup> As recommended by the *TEI Guidelines*, the same practice that is used for damaged text is also used for annotating illegibility resulting from deletion, as described below under subsections 11.6.2 and 11.6.3.

<sup>98</sup> It should be noted that the scope for the annotation of physical damage is the entire manuscript. For example water damage or staining that affects several manuscript leaves, some only marginally, is annotated only for those pages where it affects text, as this is sufficient to fulfil the twofold aim mentioned above.

text, the element is placed on the highest possible level in the hierarchy, meaning that damage affecting one or more entire words or larger textual segments is annotated by inserting the <damage> element as the parent element of the relevant elements, while damage affecting only a part of a word (or other textual container) is inserted as the child of the lowest-level element whose contents are not entirely covered by the damage.<sup>99</sup> However, in order to facilitate the processing of the text and to lower the potential for overlap, elements describing the physical aspects of the document (including not only <damage> but also <unclear> and <hi>) are *segmented* or broken into several separate elements not only at the boundaries of textual structure (to conform to the OHCO model), but also at boundaries of physical structure marked by milestone elements, such as line and page breaks. While this is not strictly necessary in terms of the XML document model, it facilitates the extraction of structural elements such as lines and pages from the edition and also makes sense from a semantic point of view, as line, column and page breaks do indicate a discontinuity in the physical structure of the document, to which also the annotation of damage pertains. The fact that several <damage> elements describe instances of damage resulting from a single source is indicated using the @group attribute, as described below.

### Cause of the damage

The attributes used to characterize the nature, cause and extent of the damage are @agent, @degree, @extent, @unit, @quantity, and @group. The @agent attribute “categorizes the cause of the damage” (TEI Consortium 2014: 725) using the following values, defined for this edition:

<b>@agent</b>	
<b>'defect'</b>	an original defect, such as uneven, hard or otherwise compromised surface in the writing medium, which has adversely affected the writing on it
<b>'hole'</b>	a penetrating hole in the writing surface, originating either before (a hole in the skin of the animal) or after (made e.g. by a bookworm) the writing of the text
<b>'loss'</b>	the loss of a part of the writing surface at its edge, indicated either by the partial loss of writing or by the leaf being of a different shape than the others
<b>'rubbing'</b>	loss of ink from the writing surface due to physical abrasion of the surface (indicative of imperfect penetration of the ink in the first place)
<b>'stain'</b>	a visible stain on the writing surface of some liquid or semi-liquid substance
<b>'tear'</b>	a tear in the writing surface, not involving any loss of material
<b>'water'</b>	discolouration of the page and bleeding or partial dissolution of the ink resulting from moisture affecting the writing surface
<b>'wrinkle'</b>	a deformation of the writing surface resulting in a sharp crease, obscuring a narrow portion of the surface

<sup>99</sup> Cases where the damage extends over multiple whole container elements and a part of an additional container are thus annotated using two separate <damage> elements, linked together using the @group attribute as described below.

## Extent of the damage

The `@degree` attribute is used for recording the approximate degree of damage on a scale from 0 (undamaged) to 1 (completely destroyed) with the accuracy of one decimal. This scale, although numerical, is of course highly subjective and should not be considered absolute in any sense. It tries to take into account both the spatial and substantial degree of coverage and should be considered in effect a product of these two aspects, the spatial degree being expressed in relation to the total content of the `<element>`, which can span anything from a single letter to an entire line of text. For example damage that covers 50 % of a letter and completely covers or obliterates it is considered to have a `@degree` value of '0.5', as is damage that covers the entire letter but reduces its visibility approximately by half. On the other hand, damage that covers 50 % of a letter and reduces its visibility approximately by half is considered to have a `@degree` value of '0.2' or '0.3'.<sup>100</sup>

The `@unit`, `@quantity`, and `@extent` attributes from the 'att.dimensions' class (TEI Consortium 2014: 733) are used to indicate the physical dimensions of the damage independently of the extent of text it affects.<sup>101</sup> The unit of measurement used in the present edition for instances of damage not affecting text is millimeters (encoded in the `@unit` attribute with the value 'mm'). The `@quantity` attribute indicates the diameter (for holes, stains etc.) or width from the edge of the page (for excisions, water damage, etc.) of the damage, while the `@extent` attribute is used to further characterise the extent of the damage in some useful way (e.g. "the edge of the page" or "a triangular shape").<sup>102</sup>

Since the effects of a single physical disturbance often extend over a large and discontinuous area, the `@group` attribute is used to group together related instances of damage, represented by individual `<damage>` elements, that are "regarded as forming part of the same physical phenomenon" (725). This attribute is given a numerical value, which is shared by each `<damage>` element belonging to the same group. Since the groups are intended to facilitate the collective description of more or less contiguous areas of damage, a damage group is here considered to consist of those `<damage>` elements describing damage resulting from a single agent and occurring on a single page. Thus damage affecting multiple pages, for example as a result of exposure to water, is numbered separately for each page.

<sup>100</sup> In terms of the alternative scale of 'high', 'medium' and 'low', also suggested by the *TEI Guidelines*, the numerical values from '0.1' to '0.3' should be seen to correspond to low, from '0.4' to '0.7' should be seen to correspond to 'medium' and from '0.8' to '1.0' should be seen to correspond to 'high'. The numerical representation, while admittedly no more accurate and objective than the keyword representation, is here adopted due to its suitability to the 'multiplicative' thinking described above. Since the horizontal coverage of the element is accurate to the level of individual letters, the spatial degree is mainly used to represent the vertical coverage of the damage, i.e. an entire line annotated with a `<damage>` element of `@type` 'loss' and `@degree` of '0.5' means that either the top or bottom half (implied by the line's location on the page) of the line has been excised.

<sup>101</sup> For damage not affecting text, at least the `@unit` and `@quantity` attributes are always used, as they are the only source of information about the extent of the damage. For damage affecting text, they are used to provide additional information about the physical extent of damage involving the loss of portions of the page.

<sup>102</sup> In the case of damage affecting text, the extent of the damage is indicated primarily by the combination of the amount of text contained within it and the `@degree` attribute described above.

### Missing folia

A special case of damage is constituted by the loss of entire folia, which not only contain an unknown amount of text, but also cut across various text-structural elements, rendering them incomplete.<sup>103</sup> The fact that pagination is indicated using milestone elements which do not have a closing tag means that the `<damage>` and `<gap>` elements cannot be placed *in between* the two pages as would be semantically appropriate.<sup>104</sup> While it would be logical to include separate `<damage>` and `<gap>` elements within the incomplete elements to indicate the missing portions, the amount of text missing from the incomplete textual items is usually unknown, since the number of lines frequently varies from page to page.<sup>105</sup> Furthermore, the *TEI Guidelines* do not allow the `<damage>` element to occur outside of a paragraph-level element, which means that it cannot be used to annotate the missing pages on its own. As a means of avoiding these problems, the present edition uses a `<note>` element of `@type 'damage'`, which can be used almost anywhere and provides a paragraph-level context to contain the `<damage>` element. Since the loss of manuscript folia usually involves the loss of text, a `<gap>` element indicating the number of lost pages (see subsection 11.6.3) is placed within the damage element. In cases where the loss of pages requires further description, a *textual note* referring to either the `<damage>` element or to the enclosed `<gap>` element is used (see section 11.8).

### 11.6.2 Unclear passages

While heavily damaged text is naturally less likely to be legible with complete confidence, different types of damage affect legibility in different ways, and not all instances of illegibility are due to physical damage to the manuscript. The necessity of indicating uncertainty and ambiguity in transcriptions of manuscript texts has been long recognised in the German editorial tradition (see e.g. Maas 1958: 17 and Zeller 1995a: 48-9), and more recently even Anglo-American editors have recognised the importance of editorial admission and documentation of uncertainty and ambiguity instead of an assured editorial establishment of the text, as well as the advantages offered by digital annotation in this regard (Taylor 2004: 98; Eaves 2006: 226; Mahoney 2006: 235-6):

The alternative to overt speculation is not neutrality but rather covert speculation and unconscious assumption, which makes no effort to seek out what evidence might be found and gives free rein to romantic stereotypes.  
(Taylor 2004: 98)

Since it is important to record the potential uncertainty in the reading of an unclear passage, the `<unclear>` element, defined by the *TEI Guidelines* for annotating

<sup>103</sup> Any textual containers rendered incomplete due to damage are separately annotated as such as described above in subsection 11.2.6.

<sup>104</sup> One possibility would be to include a 'phantom' `<pb>` element to indicate the beginning of the missing pages (or even of all of them separately), but this would mean representing pages that in fact *are not* present in the document.

<sup>105</sup> As a compromise, a `<gap>` element with a descriptive `@extent` value like "beginning of the recipe" is placed as the first or last element of the lowest-level textual container that has been left incomplete by the loss (see subsection 11.6.3 below).

parts of the text “which cannot be transcribed with certainty because it is illegible” (TEI Consortium 2014: 1472-3), is used to explicitly annotate them. In the present edition, the `<unclear>` element is used in three different contexts, reflecting three potential sources of uncertainty in the transcription. In cases where accidental damage or intentional deletion has affected the legibility of textual content, the `<unclear>` element is used within the `<damage>` or `<del>` element to indicate this, and where the uncertainty results from imperfect ductus, unclear hand or ambiguous series of minims, the element is used on its own inside any textual container element

The placement of the `<unclear>` element follows the same logic as the placement of the `<damage>` element, being placed on the highest possible level in the hierarchy, as the parent of all the elements whose entire contents are unclear and as the child of the lowest-level element containing also text that is not unclear.<sup>106</sup> Although the reason for the uncertainty of reading can in most cases be implied from the parent element (or lack thereof), the `@reason` attribute is used to indicate it explicitly using the following values, defined for this edition:

<b>@reason</b>	(on the <code>&lt;unclear&gt;</code> element)
<b>'damage'</b>	the passage cannot be transcribed with confidence due to physical damage
<b>'deletion'</b>	the passage cannot be transcribed with confidence due to being deleted
<b>'illegibility'</b>	the passage cannot be transcribed with confidence due to some irregularity or ambiguity in its letterforms

In addition to the `@reason` attribute, two attributes from the `'att.certainty'` class are used to indicate the person responsible for making the interpretation, i.e. the editor, and his certainty of the unclear reading. The `@resp` attribute records the identity of the transcriber using a pointer referring to a `<persName>` element in the header, while the `@cert` attribute indicates the certainty of the reading using the following values defined by the *TEI Guidelines* (1516), intended to allow a more fine-grained encoding of the degree of uncertainty than the traditional binary methods used for example in the ‘Leiden system’:<sup>107</sup>

<b>@cert</b>	
<b>'high'</b>	the unclear segment is still readable with a great degree of confidence and the interpretation is most likely correct
<b>'medium'</b>	the unclear segment is still readable with some degree of confidence and the interpretation is reasonably likely correct
<b>'low'</b>	the unclear segment is not readable with confidence and the interpretation is uncertain

<sup>106</sup> Again, cases where the unclear passage extends over multiple whole container elements and a part of an additional container are thus annotated using two separate `<unclear>` elements, as described for damage above (although the `@group` attribute is not used for `<unclear>`). Just as with damage, lines, columns and pages are also considered to constitute textual elements, although they are technically annotated using milestone elements.

<sup>107</sup> The Leiden system, used for papyrological and epigraphic transcription uses subpunctuation to indicate the transcribed content as uncertain. This convention of indicating readings either as certain or uncertain was also adopted by the EpiDoc adaptation of the *TEI Guidelines* (Mahoney 2006: 236), even though Terras (2010: 49–50) has pointed out that this kind of a binary representation is not very well suited to encoding a graded phenomenon like uncertainty, as it offers no way of indicating the extent of the uncertainty.

### Uncertainty of annotation

In addition to uncertainty in the transcription of textual content, the present edition also employs the @cert attribute on other elements to indicate editorial uncertainty in the evaluation or interpretation of some aspect of the annotation represented by the element. Since the present edition recognises the interpretive and thus fundamentally uncertain nature of all annotation, both descriptive and analytical, only those instances of annotation (as well as transcription) that are less certain than other instances of the same type of annotation are explicitly annotated as uncertain. This means that the absence of a @cert attribute does not mean that the interpretation represented by the annotation is absolutely certain, but rather that it is as certain as is typical to the specific type of interpretation.<sup>108</sup> In the present edition, the @cert attribute is used to indicate uncertainty mainly in connection with analytical annotation, including expansions of abbreviations (see section 11.7) and normalised forms of words (see subsection 11.9.1), where the uncertainty pertains to the content of the element on which it occurs.<sup>109</sup>

#### 11.6.3 Missing text

In addition to unclear passages, the manuscripts also contain passages of text that are either entirely unreadable or physically lost, due to either damage or intentional deletion. These passages—which cannot be transcribed—are represented in this edition using the empty <gap> element, which the *TEI Guidelines* define for marking “a point in the text where nothing at all can be read, whether because of authorial or scribal erasure, physical damage, or any other form of illegibility” (368). The <gap> element is used very similarly to the <unclear> element in that it can be placed inside a <damage> or <del> element where the illegibility is due to damage or a deletion, the only difference being that the <gap> element by definition contains no text.<sup>110</sup> Like the <unclear> element, it can also occur on its own. In addition to cases where it is used similarly to the <unclear> element to indicate the omission of content due to illegibility, it occurs on its own also in situations where the loss of folia (see above) has resulted in the loss of the end or beginning of a logical text element like a recipe (and possibly a paragraph), or a list of recipe titles. In these cases the <gap> element is placed as the first or last child element of the smallest text-structural element that is missing content and its extent is characterised using the @extent attribute (see below).

Similarly to the <unclear> element, the <gap> element takes the @reason at-

<sup>108</sup> Different types of annotation can be considered to have different levels of certainty by default: generally speaking analytical annotation such as the expansion of abbreviations can be considered to be less certain than more purely descriptive one, such as annotation of visual highlighting through rubrication. It should also be noted that the three values of the @cert attribute are all considered to signify less-than-usual certainty, even the value 'high' indicating that the editor is 'less certain than normally but still relatively certain'.

<sup>109</sup> In some cases the attribute is also used on the <gap> element, where it refers to the attribute values indicating extent of the gap, i.e. @unit and @quantity, indicating that the amount of text missing at the point is not deducible from the context with reasonable certainty, and the attribute values are merely an editorial estimate or an educated guess.

<sup>110</sup> The <gap> element with a @reason values of 'unidentified' is also used within the <reg> element (see *Normalised forms* in subsection 11.9.1) to indicate that no normalised form for the word could be supplied by the editor since the lexical identity of the word could not be established.

tribute that explicates the reason for the illegibility using the following values, defined for this edition:

@reason	(on the <gap> element)
'damage'	the passage has been lost as the result of physical damage
'deletion'	the passage has been rendered illegible as the result of a deletion
'illegibility'	the passage is illegible due to some irregularity or incompleteness of its letterforms
'abbreviation'	the word has been abbreviated and its expansion is not recoverable from the context (used only within an <expan> element; see section 11.7)
'unidentified'	while the graphemic form of the passage is clear, lexical identity can not be established (used only within a <reg> element in the in lieu of the expanded or normalised form of an unidentified word; see <i>Normalised forms</i> in subsection 11.9.1)

Since the <gap> element can represent widely varying amounts of missing textual content, the @quantity and @unit attributes are used to provide an editorial estimate of the amount of missing text. The @quantity attribute has a numerical value, while the @unit attribute provides the unit of measurement, which can be one of the following:

@unit	(on the <gap> element)
'chars'	(characters) the value of @quantity indicates the number of missing characters (defined as the space of two minims), including white space
'lines'	the value of @quantity indicates the number of missing lines of text <sup>111</sup>
'pages'	the value of @quantity indicates the number of missing pages (defined as a single side of a folio; see subsection 11.6.1)

When the <gap> element used to represent the missing content of a textual container as described above, the amount of content missing from that particular container cannot in most cases be estimated in absolute quantitative terms. In these cases the missing content is characterised by the @extent attribute, which takes as its value a short descriptive phrase, such as “end of the recipe” or “beginning of the list”.<sup>112</sup> The @resp attribute is used to indicate the party responsible for the estimation of missing text encoded using the above-mentioned attributes, and the @cert attribute to indicate that the estimation of lost textual content is unusually uncertain.

## 11.7 Annotation of abbreviation

Marking up abbreviations and their expansions is one of the most problematic aspects of the transcription of primary sources. [...] To

<sup>111</sup>In the case of gaps spanning entire lines, the <gap> element is placed on the first missing line (following a <lb> element), and the number of lines indicated by the @quantity attribute is considered to include this line.

<sup>112</sup>The absolute physical extent of the missing content is encoded in the <gap> element contained within the <damage> element representing the missing folia.



begin with, what exactly is the abbreviation? Is it the mark, sign, or letter (if there is any) that indicates that something has been suppressed, or is it the entire word? Similarly, is the expansion only the letters that have been suppressed and therefore supplied by the transcriber, or is it, again, the whole word? A case could be made for distinguishing between abbreviations with a lexical reference (suspensions, contractions, and a number of brevigraphs) and those with a graphemic reference (superscript letters and signs and the remainder of the brevigraphs). It strikes one as counterintuitive to treat the former on anything other than the whole-word level, while treating the latter in the same way seems equally misconceived.

(Driscoll 2006: 259)

As was explained in subsection 10.2.4, the present edition makes a distinction between the *abbreviation marker* that is used to indicate an abbreviation and the *abbreviated word*, of which a part has been suppressed, annotated both of them separately along with their corresponding editorial expansions and expanded word forms. Thus, unlike many editions which expand abbreviations used in the manuscript and indicate the existence of an abbreviation only by italicization (or omit any indication), this edition uses the mechanisms described by the *TEI Guidelines* (TEI Consortium 2014: 353-6) to separately encode both the original abbreviated and expanded forms of abbreviated words, making a clear distinction between what is present in the original document and what has been supplied by the editor. In addition to explicitly identifying original and editorial content, the edition also acknowledges the differing ontological status of these two aspects by encoding the original abbreviated form—which belongs to the descriptive annotation layer—in the base data file and the editorially expanded one—which belongs to the analytical annotation layer—in a separate annotation overlay linked to the base data file using the textual coordinate system.

### 11.7.1 Abbreviated forms and abbreviation markers

The original form of an abbreviated word—excluding word-final whitespace—is transcribed within an `<abbr>` element, defined by the *TEI Guidelines* for this purpose (TEI Consortium 2014: 95), placed within the `<orig>` element representing the original documentary form of the word. As mentioned in subsection 11.2.5 above, any whitespace between the abbreviated word and the following word is contained directly within the `<w>` element. Within the `<abbr>` element, the `<am>` element is used to indicate any *abbreviation markers* that have been used to signal the omission of one or more graphemes. An abbreviation marker is here defined, following the *TEI Guidelines*, as “a sequence of letters or signs present in an abbreviation which are omitted or replaced in the expanded form of the abbreviation” (TEI Consortium 2014: 354). Any special symbols apart from alphabetic characters used in the abbreviation marker are represented using a `<g>` element, as detailed under subsection 11.5.3. *XML Example 13* contains an example of the element structure used for annotating abbreviated words in the base data file.

<sup>113</sup> Since any whitespace within the `<w>` element is significant, the intra-word element structure cannot be pretty-printed without changing its content.

---

**XML Example 13:** *Representation of an abbreviated word in the transcription.*<sup>113</sup>


---

```
<w xml:id="w2210"><orig><abbr>gyng<am><g ref="#hook"/></am><abbr></orig> </w>
```

---

### 11.7.2 Editorial expansions

Since the expanded forms of abbreviated words constitute a layer of editorial analytical annotation, they are not included in the base data file but are contained in a separate annotation overlay file. This file contains a separate TEI XML document consisting of a minimal `<teiHeader>` containing only a `<fileDesc>` with a `<titleStmt>` indicating the title and author of the annotation layer, a `<publicationStmt>` indicating the authority behind its publication and its publication date, along with its distributor and availability information, and a `<sourceDesc>` describing the method through which the data contained in the overlay was produced,<sup>114</sup> followed by a `<text>` element containing the annotation data itself.

The annotation layer is associated with the appropriate base data file and the textual transcription contained in it by a `@corresp` attribute (TEI Consortium 2014: 513) whose value is a URI pointing to the `<text>` element of the appropriate base data file and an `@xml:base` attribute value that indicates the base URI to which all pointer URIs within this `<text>` element will be appended to in order to generate the full URI reference for the pointer.<sup>115</sup> The type of the annotation overlay is indicated by a `@type` value of 'annotation\_overlay' and a `@subtype` value of 'editorial', while the specific type of editorial analytical annotation is indicated by the attribute `@n`, which has a value of 'expansion'. The expanded forms for all abbreviated words are contained in the `<body>` of the text within a `<list>` element of `@type` 'wordlist' as a series of `<item>` elements, each containing a single `<w>` element. These `<w>` elements are linked to the appropriate abbreviated form by a `@corresp` value pointing to the `@xml:id` of the appropriate word in the base data file.<sup>116</sup>

Within the `<w>` element, the full expanded form is transcribed within an `<expan>` element (TEI Consortium 2014: 95) with the graphemes supplied by the editor annotated using one or more `<ex>` elements, defined by the *TEI Guidelines* for “a sequence of letters added by an editor or transcriber when expanding an abbreviation” (354). Expanded forms that the editor considers uncertain are marked by the `@cert` attribute with a value of 'high', 'med' or 'low', depending on his confidence in the expansion.<sup>117</sup> Abbreviated words for which no expanded form can be established are annotated with an `<expan>` element containing a `<gap>` element with a `@reason` attribute value of 'abbreviation'. An example of the internal structure of the annotation overlay is shown in *XML Example 14*.

---

<sup>114</sup> The internal composition of these documentary elements is similar to that described in section 11.1 for the base data file itself.

<sup>115</sup> In the present edition, relative file paths are used for the URI in order to allow linking between the different files irrespective of the absolute location of the edition source files on the user's system. In a future version deployed over the Internet, these will be replaced by persistent URLs using the HyperText Transfer Protocol scheme.

<sup>116</sup> Expansions of partial words also replicate the `@part` attribute of the original `<w>` element in the base file to indicate their incomplete status.

<sup>117</sup> Since the expansion of abbreviations is an entirely conjectural operation even under the best of

---

**XML Example 14:** *The internal structure of the <text> element of the annotation overlay containing the expanded forms of the abbreviated words contained in the base data file for a single manuscript.*

---

```
<text type="annotation_overlay" subtype="expanded_abbreviations"
  corresp="MSHarley4016.xml#Harley4016" xml:base="MSHarley4016.xml">
  <body>
    <note type="wordlist">
      <w corresp="#w10">
        <expan>m<ex>iste</ex>r</expan>
      </w>
      ...
      <w corresp="#w21007">
        <expan>Jan<ex>uary</ex></expan>
      </w>
    </note>
  </body>
</text>
```

---

### 11.7.3 Relating expansions to abbreviated words

This annotation, which is redundant in the sense that it repeats each abbreviated word in its entirety, has been selected because it allows for the separate but linked representation of both the original orthography of the word and its editorial interpretation, complete with information about the status of each character in the word—whether it is original or supplied by the editor. In cases where there is a one-to-one correspondence between an abbreviation marker and an expansion, it also allows for their linking, either implicitly by their order within their respective parent elements or explicitly using the @n attribute where there is room for ambiguity, as in cases where not all <am> elements have a corresponding <ex> element or they occur out of order.<sup>118</sup> In the case of abbreviated words (whether indicated by an abbreviation marker or not) for which no expanded form can be deduced—mainly names and other words abbreviated by suspending all letters except for the first one—a <gap> element with a @reason attribute value of 'abbreviation' is placed within the <expan> element. Expansions which are considered to be unusually uncertain for some reason are annotated by using the @cert attribute with an appropriate value (see subsection 11.6.2) on the <expan> element.<sup>119</sup>

Words that have been divided into two by a line break (see subsection 11.2.6) are treated as two separate words in terms of annotating the abbreviation and its expansion.<sup>120</sup> In cases where both of the halves of the word are involved in the

---

circumstances, the threshold for explicitly annotating an expansion as uncertain is relatively high.

<sup>118</sup> In this edition, the @n attribute has been used only in these cases. Whenever there are at least as many <am> elements as there are <ex> elements, their orders should be understood to correspond to each other (i.e. the last <am> elements will be left without a corresponding <ex> element). In cases, where there is one <am> element and several <ex> elements with no explicit @n attributes, all of them should be understood to correspond to the one <am> element. All <am> elements that have no corresponding <ex> element should be understood as representing otiose markers, while <am> elements that correspond to several <ex> elements should be understood as *general* markers of abbreviation (Hector 1966: 29).

<sup>119</sup> It should be noted that since the expansions of abbreviations are by default conjectural and thus uncertain, the threshold of explicitly annotation an expansion as uncertain is relatively high.

<sup>120</sup> This means that it is possible—and in fact quite common—for one half of the word to be abbreviated and the other unabbreviated. In these cases, any processing algorithm used to combine separated words must ensure that the combined word is marked as abbreviated and appropriately annotated

abbreviation, the graphemes supplied in expanding the word are divided between the two parts of the word so that those graphemes added to any point preceding the first grapheme of the second half are included in the first half (contained within one or more <ex> elements), and those graphemes added to any point following it are included in the second half.<sup>121</sup>

## 11.8 Editorial notes

While all of the annotation described in this chapter—whether documentary, descriptive and analytical—can be seen as a form of ‘editorial notes’ in the wider sense, the title of this section refers to editorial notes in the more traditional sense of written remarks associated with some point or span in the text and providing additional information about it. Although most of the features of the original manuscript can be adequately described using the kinds of formal annotation described in this chapter, informal prose description has been used to further describe certain aspects of the work, the text, and the document. These *editorial notes* are of two kinds: *textual notes* which describe physical and textual features present in the manuscript and thus belong to the *descriptive* layer, and *explanatory notes*, which are used to explicate the meaning of the text in places where it is unclear and to comment on the wider significance of a certain passage, being thus clearly analytical.<sup>122</sup>

The fact that these two types of notes belong to different annotation layers and operate on different levels of the textual object also has implications to their scope; while textual notes always apply to a specific point in a specific manuscript document, explanatory notes frequently apply to a point in the *work* or some versions of it, and thus to one or more points in one or more documents. This, in turn, has implications for their encoding and storage. Although both of them are separated from the transcription itself by using what is known as *reverse linking* (Guralnik 2002), the descriptive textual notes which are more closely integrated with the annotated transcription of a single and refer not only to the textual content but also to various other textual phenomena—emendations, decoration, damage etc.—are included in the metadata header of each base data file, while the explanatory annotation—by definition concerned only with the abstract textual content—is stored in a separate file constituting a separate editorial *annotation overlay*, and linked either to entire recipes or to specific spans of textual content within the six base data files.

---

if *either* of the halves are abbreviated.

<sup>121</sup> In some rare cases this may mean that the <am> and <ex> elements corresponding to each other are contained in different parts of the word, so this will need to be taken into account in processing the text if explicit association between abbreviation markers and their expansions is sought.

<sup>122</sup> This distinction between textual and informative notes was introduced already in the 1940s by Julian Boyd, the first editor of *The Papers of Thomas Jefferson* series, but despite the fundamental ontological difference between them, few documentary editors seem to have followed his example (Kline and Holbrook Perdue 2008: 242).

### 11.8.1 Textual notes

While textual notes belong to the descriptive annotation layer and thus constitute a part of the base data file of the edition, they represent textual content not present in the original document and should therefore be kept apart from the transcription. For this reason, textual notes are encoded as `<note>` elements of `@type` 'textnote', stored within within the `<notesStmt>` element of the header (see subsection 11.1.2) and linked to the appropriate point or span of text in the transcription by a `@target` URI pointing to the `@xml:id` of an element representing the textual or documentary phenomenon the note refers to. In cases where no suitable text-structural element exists, a semantically neutral `<seg>` element with an `@xml:id` has been added to represent the referential target of the note. In addition to the `@type` and `@target` attributes, each note also has its own `@xml:id` in order to allow unambiguous reference to it both by further annotation overlays and scholarly discussions of the edition. Similarly to other interpretive elements, the author of the note—usually the editor—is indicated by a `@resp` attribute pointing to an entity defined in the header.

Notes which are associated with an element like `<damage>` or `<add>` and contain essentially the same information—usually with additional detail—as the formal representation in the element itself also have an `@exclude` attribute pointing to the element, signalling that the textual note replicates the formally expressed information in prose form and should be used instead of any notes that would otherwise be automatically generated from the formal representation.<sup>123</sup> *XML Example 15* shows an example of a textual note with the attributes used for identifying it and linking it to the text.

---

**XML Example 15:** *Textual note and its location in the header.*

---

```
<TEI>
  <teiHeader>
    ...
    <notesStmt>
      ...
      <note xml:id="textnote_4" type="textual_note" target="#dam212 #dam211"
        resp="#VM" exclude="#dam212 #dam211">
        <p>There is a dark brown stain of unknown substance, c. 8 mm in
          diameter, at the fold of ff. 10v and 11r.</p>
      </note>
      ...
    </notesStmt>
    ...
  </teiHeader>
  <text>
    <!-- The transcription itself -->
  </text>
</TEI>
```

---

<sup>123</sup> In the diplomatic transcription included in the present edition, automatic footnotes are generated for instances of damage, scribal emendation, illegibility and various other documentary features. Textual notes associated with representations of these features that do not have an `@exclude` attribute are appended to these automatically generated notes and have been formulated accordingly, while ones which feature the attribute will suppress and replace the automatically generated note.

### 11.8.2 Explanatory notes

Explanatory notes, on the other hand, have been implemented entirely separately from the base data file in order to allow their linking to several manuscript versions and to demonstrate the concept of external *annotation overlays*. Thus, instead of a separate file for each base data file as in the case of the editorial expansions described in subsection 11.7.2, all of the explanatory notes pertaining to any version of the *Potage Dyvers* collection are stored in a single annotation overlay file (PD\_explanatory\_notes.xml). The header of the annotation overlay file contains the same elements as that described for editorial expansions, with the addition of a <listBibl> element to the <sourceDesc>, containing structured bibliographical entries for all the sources (primary and secondary) referred to in the explanatory notes.

The <text> element of the annotation overlay file uses the same @type and @subtype values as the one described in subsection 11.7.2 for editorial expansions to identify its contents as an editorial annotation overlay, and an @n value of 'explanatory\_notes' to identify its specific nature. As described for expansions, this annotation overlay also uses the @corresp attribute to associate itself with the appropriate base data files, in this case listing pointers to all of the six PD versions, separated by whitespace.<sup>124</sup> The content of the <text> element is made up of two different components: linking metadata contained within its <front> and the textual content of the notes, contained within the <body>.

#### Linking explanatory notes

The <front> of the annotation overlay contains metadata used to associate the explanatory notes either to specific recipes or spans of text in the six manuscript versions of the *Potage Dyvers* collection. This metadata is expressed in the form of <ptr> and <span> elements, defined by the *TEI Guidelines* (TEI Consortium 2014: 96-7, 550-4) for this purpose. Of these, the former is used to associate notes with all of the parallel versions of an individual recipe by linking to the PD\_recipe\_versions.xml (see subsection 10.3.2), while the latter is used to associate notes with individual spans of text (down to an individual word) within one or more base data files. The function of these elements is to serve as 'intermediate pointers', and to define virtual entities—either groups of recipes or spans of consecutive word units—to which the notes themselves can refer to.

The two units differ from each other both syntactically and semantically. While the <ptr> elements are understood purely as unidirectional one-to-many links that define a link between an element pointing to them and the elements they point to without implying the aggregation of their target elements, the <span> elements are understood to define a new aggregate entity, a virtual element in the target base data file, containing the elements explicitly indicated by its @from and @to attributes and all the elements located between them.<sup>125</sup> Any note linked to the

<sup>124</sup> Because of the multiple targets to which this annotation overlay refers, the @xml:base attribute is not used on the <text> element but rather on the separate <spanGrp> elements defined for each base data file (see below).

<sup>125</sup> The conceptualisation of the span as a 'virtual element' which can be replaced by an actual one is intended to simplify the association of the notes to the base data file by shifting the responsibility of maintaining well-formedness from the processing stage to the encoding stage. Essentially this

<span> element should be understood as being associated with the contents of this virtual aggregate element.<sup>126</sup> In other words, a <span> element can be replaced by a new element in the target document, while the <ptr> element cannot. This also means that while the <span> elements are always associated with a single base data file, the <ptr> elements usually refer to several base data files.

This difference is also visible in their organisation within the <front> element, exemplified in *XML Example 16*; while all of the <ptr> elements are located within a single <linkGrp> element of @type 'recipese', identified by the @xml:id 'PD\_recipese' and associated with the PD\_recipe\_versions.xml file using the @xml:base attribute, the <span> elements are grouped into six separate <spanGrp> elements of @type 'wordspan', each of which is identified by its own @xml:id and associated with a single base data file (again using the @xml:base attribute). Within the <linkGrp> element, each of the <ptr> elements is identified by its own unique @xml:id (here of the form 'PD\_en\_xx') and contains a @target attribute whose value is a fragment identifier referring to a <link> element within the PD\_recipe\_versions.xml representing a single unique recipe. This <link> element in turn refers to all of the individual versions of that recipe in the various base data files, as described below in subsection 11.9.2, linking the explanatory note(s) referencing the <ptr> element to all of the individual recipe versions. The <span> elements are similarly identified by an @xml:id, but instead of pointing to an external file, they define a span of consecutive word-units in the base data file indicated by the @xml:base attribute of the containing <spanGrp> by indicating the first and last elements to be included in the span using the @from and @to attributes.<sup>127</sup>

### Encoding explanatory notes

As the apparatus for linking the explanatory notes to the base data file is contained separately in the <front> part of the annotation overlay, the encoding of the explanatory notes themselves is relatively straightforward. The notes are contained within a <div> element of @type 'notes' inside the <body> of the overlay file, each explanatory note being represented by a <note> element with a unique @xml:id, a @type of 'explanatory\_note' and a @resp attribute indicating the party responsible for the contents of the note (here the editor).<sup>128</sup> The span of textual content or set of recipe versions to which the note refers is indicated by the @target attribute, whose value consists of one or more pointers to the relevant <ptr> or <span> elements—usually only one or the other in any given note—defined in the <front> of the overlay file. The fact that the target of reference is not the intermediate link but rather the last element in a chain of pointers is indicated by an

---

requires that each span constitute a well-formed XML element, i.e. that the elements pointed to by the @from and @to elements be siblings of each other.

<sup>126</sup> It should be noted that unlike in the examples presented in the *TEI Guidelines* (TEI Consortium 2014: 550-4) where the <span> element itself contains the analytical data associated with the span defined by the @from and @to attributes, the <span> element is here used in a purely intermediary role with no semantic content of its own.

<sup>127</sup> In cases where the span consists of a single word, no @to element is provided, as per the recommendation in the *TEI Guidelines* (550).

<sup>128</sup> The intervening <div> element is required by the TEI content model but serves not other purpose. The inability of the <body> of a document to contain a series of <note> elements without an intervening <div> is a completely arbitrary restriction of the TEI content model, and is most likely the result of mere oversight.

---

**XML Example 16:** *The structure of the apparatus for linking explanatory notes to sets of recipes and spans of word units, located in the <front> of the overlay file.*

---

```

<TEI>
  <teiHeader>
    ...
  </teiHeader>
  <text type="annotation_overlay" subtype="editorial" n="explanatory_notes"
    corresp="MSAdditional5467.xml#Additional5467 MSAshmole1439.xml#Ashmole1439
      MSCosinViii11A.xml#CosinViii11A MSDouce55.xml#Douce55 MSHarley279.xml#Harley279
      MSHarley4016.xml#Harley4016">

    <front>
      <linkGrp type="recipeset" xml:id="PD_recipesets">
        <ptr xml:id="PD_1" target="MSCosinViii11A.xml#CosinViii11A_r1
          MSAdditional5467.xml#Additional5467_r1
          MSAshmole1439.xml#Ashmole1439_r1
          MSDouce55.xml#Douce55_r1
          MSHarley279.xml#Harley279_r1"></ptr>

        ...
      </linkGrp>
      ...
      <spanGrp type="wordspan" xml:id="CosinViii11A_wordspans" xml:base="MSCosinViii11A.xml">
        <span xml:id="CosinViii11A_en_1" from="#w10405a" to="#w10407"></span>
        <span xml:id="CosinViii11A_en_2" from="#w14174"></span>
        ...
      </spanGrp>
      ...
    </front>
    <body>
      ...
    </body>
  </text>
</TEI>

```

---

@evaluate attribute with the value 'all' (TEI Consortium 2014: 758). The content of the note is contained in one or more <p> elements within the <note> element and consists of text annotated with the elements defined for the annotation of meta-data in subsection 11.1.1. For all bibliographical references contained in the notes, a bibliographical entry describing the cited work is included in the <sourceDesc> element of the TEI header. *XML Example 17* shows an example of an explanatory note.

## 11.9 ‘Non-editorial’ analytical annotation

In addition to the descriptive and analytical annotation produced as a part of the editorial process, the present edition also incorporates two different types of annotation that are not considered to be editorial but rather the result of further scholarship taking place after it. As described in section 10.3, these include the linguistic annotation of the normalised form, basic word class and linguistic identity of each lexical unit in each of the *PD* versions, and the annotation of the intertextual relationships between *parallel versions* of the same recipe occurring in different versions of the collection.



---

**XML Example 17:** *The structure of explanatory notes, located within the <body> of the overlay file.*

---

```

<TEI>
  <teiHeader>
    ...
  </teiHeader>
  <text type="annotation_overlay" subtype="editorial" n="explanatory_notes"
    corresp="MSAdditional5467.xml#Additional5467 MSAshmole1439.xml#Ashmole1439
      MSCosinViii11A.xml#CosinViii11A MSDouce55.xml#Douce55
      MSHarley279.xml#Harley279 MSHarley4016.xml#Harley4016 ">

    <front>
      ..
    </front>
    <body>
      ...
      <note xml:id="expnote_13" type="explanatory_note" target="#CosinViii11A_en_4" resp="#VM">
        <p><ptr type="biblref" rend="textcite" target="#Heatt_and_Butler_1985" n="181"></ptr>
          suspect this unknown word to be a miscopying of <mentioned>lounge</mentioned>
          (<gloss>tongue</gloss>, from Fr. <mentioned>langue</mentioned>), the item referring
          to a dish of <soCalled>corned</soCalled> (i.e. salted) tongue.</p>
      </note>
      ...
      <note xml:id="expnote_18" type="explanatory_note" target="#Douce55_en_6" resp="#VM">
        <p>The <quote>J. Chicheley</quote> who is here identified as the person from whom
          Urry borrowed this manuscript is most likely <persName><forename>John</forename>
          <surname>Chicheley</surname></persName>, barrister-at-law in the Middle Temple
          from <date when="1701">1701</date>.</p>
      </note>
      ...
    </body>
  </text>
</TEI>

```

---

### 11.9.1 Linguistic annotation

While many corpus-linguistic methods are hindered by the presence of orthographic variation in historical texts (see sections 4.1 and 5.6), they simultaneously depend on it as an important component of their primary data. In order to overcome this problem, the present edition supplements the faithful graphemic transcription of each separately annotated morphosyntactic word, i.e. the linguistic data, with metadata consisting of its normalised form, basic word class and linguistic identity (Middle English or foreign), according to the principles described in subsection 10.3.1. Since this linguistic annotation constitutes a layer of analytical annotation beyond the annotation of textual structure, it is not included in the base data file but contained in a separate *annotation overlay* linked to the original orthographic forms in the transcription using the *textual coordinate system* formed by the uniquely identified word units annotated in the base data file (see *Word-units as textual coordinates* in subsection 10.1.3), preserving the integrity of the data itself.

Like the expansion overlay described in subsection 11.7.2, this file contains a minimal <teiHeader> along with a <text> element containing the annotation data itself, associated with the appropriate base data file using the @corresp and @xml:base attributes. The fact that the overlay contains linguistic analytical annotation is indicated by a @type value of 'annotation\_overlay' and a @subtype value of 'linguistic', while the specific nature of linguistic annotation—namely a combination of *normalisation* and *word classification*—is indicated by the @n at-

tribute value of 'normalisation+wordclass'. If the base data file contains words which have been segmented using the @part attribute (values 'l' and 'f'), a <joinGrp> element with the @result value of 'w' (TEI Consortium 2014: 528), containing a <join> element for each such word is located within the <front> of the <text> element. These elements have two attributes, @xml:id and @target, the latter pointing to the @xml:id values of the two halves, while the value of the former is the shared initial part of these @xml:id values without the *a* or *b* suffix. The normalised forms and word classes of all morphosyntactic words in the base data file, i.e. the metadata itself, are contained in the <body> of the text within a <list> element of @type 'wordlist' as a series of <item> elements, each containing a single <w> element linked to the appropriate original form by a @corresp value pointing to the @xml:id of the appropriate word in the base data file.

### Normalised forms

For each of these <w> elements, a *normalised form*, chosen according to the principles outlined in subsection 10.3.1, is encoded as the value of a <reg> element, defined by the TEI Guidelines (TEI Consortium 2014: 81-2) for the purpose of annotating normalised spellings of non-standard word-forms, placed within the <w> element. The authority responsible for the identification of the word and its association with the appropriate normalised form is indicated using the @resp attribute, referring to one or more persons identified in the header. The dictionary or other source used as the basis of the normalisation is indicated using the @source attribute containing a pointer to a bibliographic entry in the <sourceDesc> of the header. Cases where there is uncertainty about the correctness of the normalised form are marked by using the @cert attribute with a value of 'high', 'medium' or 'low' on the <reg> element to reflect the confidence of the responsible party in the normalisation.<sup>129</sup> The normalised form of a word whose lexical identity cannot be established and for which no normalised form can thus be provided, is replaced by a <gap> element with a @reason attribute value of 'unidentified'.

Words that are broken up by line breaks and annotated as two <w> elements in the base data file are represented by a single <w> element in the linguistic annotation overlay, as the metadata represented by it is not mappable between the two halves of the word but pertains equally to both halves. The value of the @corresp attribute in these cases is the @xml:id of the original form without the appended letter suffix (which is shared by both halves of the word), pointing to a <join> element in the <front>, which in turn links the normalised form to both halves of the original form.<sup>130</sup> Words whose original forms are incomplete (@part value of 'Y') are annotated with the full normalised form of the word, if it can be established (using the @cert attribute to encode any possible uncertainty). The structures used for annotating the normalised form, along with the word class and language of the word—which are described below—are shown in *XML Example 18*.

<sup>129</sup> It should be noted that full confidence in the regularization is indicated by the absence of the @cert attribute, and its presence thus always implies a level of confidence lower than the norm, regardless of its value.

<sup>130</sup> To indicate that the target of the @corresp reference is the <join> element within the <front> of the annotation overlay itself, all such words also have an empty @xml:base attribute.

---

**XML Example 18:** *An example of the structure of the annotation overlay file for linguistic annotation.*

---

```

<TEI>
  <teiHeader>
    ...
    <notesStmt>
      <note type="classification">
        <interpGrp type="wordclass">
          <interp xml:id="N">noun</interp>
          <interp xml:id="P">pronoun</interp>
          <interp xml:id="V">verb</interp>
          <interp xml:id="J">adjective</interp>
          <interp xml:id="AV">adverb</interp>
          <interp xml:id="PP">preposition</interp>
          <interp xml:id="C">conjunction</interp>
          <interp xml:id="D">determiner</interp>
          <interp xml:id="INT">interjection</interp>
          <interp xml:id="U">unknown</interp>
        </interpGrp>
      </note>
    </notesStmt>
    ...
  </teiHeader>
  <text type="annotation_overlay" subtype="linguistic" n="normalisation+wordclass"
    corresp="MSHarley4016.xml#Harley4016" xml:base="MSHarley4016.xml">
    <front>
      <joinGrp result="w">
        ...
        <join xml:id="w589" target="#w589a #w589b"/></join>
        ...
      </joinGrp>
    </front>
    <body>
      <note type="wordlist">
        <w corresp="#w8" ana="#V" xml:lang="eng">
          <reg resp="#VM" source="#OED">bought</reg>
        </w>
        <w corresp="#w9" ana="#PP" xml:lang="eng">
          <reg resp="#VM" source="#OED">in</reg>
        </w>
        ...
        <w corresp="#w587" ana="#V" xml:lang="enm">
          <reg resp="#VM" source="#OED">kneeling</reg>
        </w>
        <w corresp="#w588" ana="#PP" xml:lang="lat">
          <reg resp="#VM" source="#OLD">in</reg>
        </w>
        <w corresp="#w589" ana="#N" xml:lang="lat" xml:base="">
          <reg resp="#VM" source="#OLD">pontificalibus</reg>
        </w>
        ...
        <w corresp="#w10002" ana="#D" xml:lang="enm">
          <reg resp="#VM" source="#OED">the</reg>
        </w>
        <w corresp="#w10003" ana="#J" xml:lang="enm">
          <reg resp="#VM" source="#OED">raw</reg>
        </w>
        ...
        <w corresp="#w21006" ana="#U" xml:lang="und">
          <reg resp="#VM"><gap reason="#unidentified"/></reg>
        </w>
        <w corresp="#w21007" ana="#N" xml:lang="eng">
          <reg resp="#VM" source="#OED">January</reg>
        </w>
      </note>
    </body>
  </text>
</TEI>

```

---

## Word classes

Being a part of the same process of lexical disambiguation as the normalised form, the basic word class is also indicated within the same linguistic annotation overlay as the normalised form. The word class is annotated using the `@ana` attribute, defined in the *TEI Guidelines* for indicating “one or more elements containing interpretations of the element on which the `@ana` attribute appears” (TEI Consortium 2014: 550). The attribute is placed on the relevant `<w>` element in the overlay file, and its value refers to an *interpretation* (550-4) defined by an `<interp>` element located within an `<interpGrp>` (interpretation group) element (550) located within a `<note>` of `@type 'classification'` in the `<notesStmt>` part of the `<teiHeader>` of the overlay file. Any uncertainty in the assignment of a word class according to the principles outlined in subsection 10.3.1 is usually associated with a more general uncertainty as to the identity of the word, and is thus factored into the general certainty encoded in the `@cert` attribute of the `<reg>` element described above. The word classes represented by the `<interp>` elements have been defined on the grounds outlined in subsection 10.3.1 and include:

<code>@xml:id</code>	(on the <code>&lt;interp&gt;</code> element)
'N'	noun
'P'	pronoun
'V'	verb
'J'	adjective
'AV'	adverb
'PP'	preposition
'C'	conjunction
'D'	determiner
'INT'	interjection
'U'	unknown

## Foreign-language words

The third dimension of linguistic annotation included in the linguistic annotation overlay is the annotation of foreign-language words occurring in the text. This aspect of linguistic annotation is a direct by-product of the normalisation process, since the normalisation of a word requires and implies also the establishment of its linguistic identity, as was explained in subsection 10.3.1. This detailed word-level annotation of language is simultaneously independent of the text-structural annotation of language on the level of text divisions and ‘chunks’, and works in conjunction with it. The linguistic annotation of language is independent in the sense that it is based entirely on lexical analysis of individual words, without regard to their larger linguistic context encoded in the textual structure. However, because of this very independence, the word-level data can be related to the chunk-level annotation in order to identify words whose linguistic identity differs from that of their linguistic context and which can thus help identify *loci* of *code-switching* (see subsection 6.3.2).

In order to distinguish Middle English words from foreign words that have not yet been acclimatised to English and to place the normalised form into its correct linguistic context, the linguistic identity of all words is annotated by an `<xml:lang>` attribute placed on each `<w>` element in the linguistic annotation overlay, with

one of the following values, defined by the ISO 639-3 standard (SIL International 2011):

@xml:lang	(on the normalised <w> element)
'enm'	Middle English (1100–1500)
'frm'	Middle French (ca. 1400–1600)
'xno'	Anglo-Norman
'lat'	Latin
'eng'	Modern English
'und'	Undetermined

11.9.2 Intertextual relationships

In addition to the layer of linguistic annotation described above, the present edition also includes another layer of analytical annotation that is not considered to belong to the process of editing the text, but rather to the macro-level textual and structural analysis of the discourse colony. This layer, presented as a separate annotation overlay, consists of links joining together all of the *parallel versions*—as defined in subsection 10.3.2—of the 371 unique recipes contained in the entire *Potage Dyvers* family. Like the overlay for editorial annotation, this overlay is by definition not tied to a single base data file or MS version but rather serves to integrate all of them into a single rhizomatic network consisting of the parallel relationships between different versions of the ‘same’ recipes and their different sequential organisation in each of the collections.

The annotation overlay containing the linking metadata is structured very similarly to the ones containing the editorial expansions of abbreviated words and the linguistic annotation, the differences being limited mainly to the <body> of the document.<sup>131</sup> The <teiHeader> of this overlay contains the same elements as the ones described above in subsection 11.7.2 for editorial expansions of abbreviations and editorial notes, providing the basic bibliographic data for the overlay.

The <text> of the overlay is identified as an overlay of intertextual annotation by the @type and @subtype values of 'annotation\_overlay' and 'intertextual', with its specific nature indicated by the @n value of 'recipe\_versions'. Similarly to the overlay for editorial notes, this one also uses the @corresp attribute to associate itself with all of the six *PD* versions and omits the @xml:base attribute. The principal contents of this overlay are contained within its <body> and consist of a series of <link> elements, defined by the *TEI Guidelines* for defining “an association or hypertextual link among elements or passages” (TEI Consortium 2014: 491). These links are grouped together by a <linkGrp> element of @type 'parallel', indicating that the items linked together should be understood as parallel versions of each other. In addition to this primary group of links, the <front> of the document also contains a second <linkGrp> with a @type of 'join', containing intermediate links used to reconstitute recipes which have been divided into several independent recipes so that they can be linked with their parallel versions in other MS ver-

<sup>131</sup> While this thesis does not yet define a formal general model for the encoding of annotation overlays, the overlays contained in the present edition are intended to serve as initial sketches for such a model, to be developed in concert with a more general set of guidelines for producing corpus linguistic editions.

sions. In order to conform to the content model of a TEI, the <linkGrp> element within the <body> is enclosed within a <div> element of @type 'links'.<sup>132</sup>

Each of these <link> elements represents a single unique recipe on the level of an abstract *work*, represented by a number of *versions* in one or more of the six manuscripts edited here. These abstract recipes are identified by a *recipe reference* encoded as the value of the @xml:id attribute, being of the form 'PD\_nn', where *nn* is a numerical identifier given to the recipe within the family. Since the ordering of any of the collections cannot be considered to be more 'authoritative' than that of the others, this numerical identifier of the recipe reference is somewhat arbitrary and is intended merely as a way of referring to individual recipes on the level of the work.<sup>133</sup>

In addition to the @xml:id, each of the <link> elements contains a @target attribute which contains a whitespace-separated list of relative URIs pointing to all those <text> elements in the six base data files that represent recipes considered to be versions of the particular recipe. While the vast majority of the recipes occur only once in each collection, there are some cases where a recipe has an *internal parallel* within the same collection. These internal parallels are indicated in the same way as all the other parallel versions, which means that the @target attribute of some recipes may contain multiple pointers to the same base data file. While the number of MSS in which a recipe occurs can be deduced from the number of unique file references within the @target attribute, this information is also explicitly indicated by an @n attribute on each link. The intermediate links—mentioned above and encoded within a separate <linkGrp> element in the <front> of the document—are used to allow linking between the parallel versions of recipes which have been divided into several individual recipes in one or more of the manuscripts. For this purpose, each of these intermediate <link> elements contains an @xml:id attribute which is used as an aggregate reference for all the parts of the divided recipe, and a @target attribute which contains pointers to those <text> elements representing the parts of the divided recipe.

As was pointed out in subsection 10.3.2, this annotation layer is based on a detailed micro-level comparison of the textual content of all the recipes and the observations made by earlier scholars, and is used in chapter 13 for the detailed macro-level analysis of the structural relationships between the six *PD* versions, as well as for the creation of both the parallel reading edition of all recipe versions and the interactive HTML edition included on the accompanying CD-ROM (appendices C and D). In addition to identifying the parallel versions within the *Potage Dyvers* family itself and allowing them to be examined and analysed in parallel,

<sup>132</sup> The inability of the <body> of a document to contain a <linkGroup> without an intervening <div> is a completely arbitrary restriction of the TEI content model, and is most likely the result of mere oversight.

<sup>133</sup> The order of numbering the recipes has been determined by first numbering the recipes shared by at least half (i.e. three) of the collections according to the order in which they occur in the collection containing the largest number of these shared recipes, namely MS D (which contains all but one of the recipes occurring in at least three collections). The rest of the recipes are numbered in the order of their prevalence (i.e. those occurring only once coming last). All of the recipes that occur in two collections are found in MSS As and H279, and occur mostly in the same order. Where their order differs, the order of MS As has been given primacy. The recipes occurring in only one version are numbered one collection at a time, in the following order: MS D, MS Ad, MS C, MS H4016, MS As and MS H279.

these recipe references can also be used as a convenient intermediate or aggregate identifier for linking the *Potage Dyvers* recipes with their parallel versions found in other collections that are edited in the future.

### 11.9.3 Textual sources of the recipes

In addition to grouping together the parallel versions of each unique recipe contained in the six *PD* collections, the present edition also contains a further 'second degree' annotation overlay that groups the unique recipes identified by the intertextual annotation according to their likely textual sources, based on the structural analysis of the six manuscript versions and the occurrence patterns of the individual recipes within them. This analysis, described in chapter 13, would seem to indicate five distinguishable sources for the recipes constituting the six collections edited here. The recipes inherited from each of these sources are grouped together by using a similar annotation overlay as that used for the intertextual linking of parallel versions, with the difference that instead of linking individual recipe versions to 371 recipe *types*, this overlay links those recipe types into six groups. This is accomplished through six `<link>` elements, the `@target` attributes of which refer to the appropriate `<link>` elements in the intertextual annotation layer described above.<sup>134</sup>

### 11.9.4 Canonical titles

In order to provide descriptive titles to the unique recipes identified in the *PD* collections, the present edition also adds a further annotation overlay linking a *canonical title*, based on the *Concordance of English Recipes* by Hieatt and Nutter (2006), to each unique recipe based on its recipe reference. While the canonical titles for the vast majority of the recipes are taken directly from Hieatt and Nutter (2006), there are some exceptions. These include the few recipes that do not occur in the edited recipe collections included in the concordance, recipes which have been mistakenly associated with a similar but unrelated title in the concordance, and recipes for which the only lemmatised form given in the concordance is of the type *headword*, *modifier*. For the first two cases, an original canonical title has been constructed analogously to those given in the concordance, while for the third, the comma-separated title has been reformatted into a simple noun phrase of the format *modifier headword*. In the latter case, the original form used in the concordance has been included as an index entry for ease of reference.

In terms of its document structure, this annotation overlay is very similar to the one containing the intertextual links, with a `<teiHeader>` that has the same components, differing only in their content. The `<text>` of the document is identified as an annotation overlay by the `@type`, `@subtype` and `@n` values of 'annotation\_overlay', 'intertextual' and 'canonical\_titles', and linked to the intertextual linking layer by the `@corresp` and `@xml:base` values of 'PD\_recipe\_versions.xml'. The titles themselves are presented within a `<list>` structure within the `<body>` of the text, each title being represented by a `<title>` element of `@type` 'canonical'

<sup>134</sup> Of the six `<link>` elements, five represent the distinct sources identified in chapter 13, while the sixth links together recipes whose origin remains unclear on the basis of its distribution in the MSS.

within a plain list `<item>`. Each title is linked to the appropriate unique recipe by a `@corresp` attribute value referencing the `@xml:id` of the appropriate `<link>` element in the intertextual annotation overlay. The canonical title as a noun phrase is contained directly within the `<title>` element while the inverted, comma-separated forms used in the concordance and mentioned above are included within a `<index>/<term>` structure following the title.<sup>135</sup>

## 11.10 Conclusion

In accordance with the editorial model for a digital documentary edition outlined in chapters 4 and 5, the annotation described in this chapter has been designed to allow for the modelling of not only the textual content of the original document, but also of its textual and physical structure, as well as those visual and physical features which have been considered as relevant to its interpretation as a communicative event. While the annotation on the level of an actual XML document has not been strictly divided into the three separate levels of *documentary*, *descriptive*, and *analytical* annotation—the main part of the base data file containing both descriptive and analytical annotation—the annotation scheme has nevertheless been designed in a way that it is possible to selectively remove either all analytical annotation or all descriptive annotation.<sup>136</sup> The *textual coordinate system* built into the annotation scheme also allows for the unlimited addition of new analytical annotation layers by the means of *annotation overlays*, whose structure and operating principle is here demonstrated by the overlays for editorial annotation of expanded forms for abbreviated words and explanatory notes, and the ‘non-editorial’ annotation of linguistic tokens and intertextual relationships.

Although the annotation system also includes a provision for non-machine processable editorial notes for the detailed description of textual and paratextual phenomena (subsection 11.8.1 above), the principal method of annotating these phenomena—including the ones further clarified by prose description—is by the use of a formal markup language, which has been designed to allow the sufficiently detailed description of even special cases using generic structures that can be processed automatically. This has been given a high priority, since as Gabler (2006) has pointed out, the automatic processing of data which constitutes one of the main strengths of digital editions requires that “the commonsense answer

<sup>135</sup> It should be noted that the canonical—or “lemmatized”, to use their own terminology—titles provided by the *Concordance* are essentially names of individual *dishes* rather than names for individual recipes, clearly different recipes describing what could be considered the same dish being as a rule given the same canonical title. This means that they can as such be used as a means of grouping together distinct recipes describing the same dish. However, while the concordance of Hieatt and Nutter (2006) does provide a means of associating each recipe—understood here on the level of an abstract *work*—to what is essentially a specific dish, the fact that no formal identifiers are provided by Hieatt and Nutter (2006) for the lemmatised or canonical forms means that this linking will for the time being remain implicit, relying on the identical textual content of the `<title>` elements. In the future—copyright considerations allowing—a more robust solution would be to create an expandable database of canonical titles, in which each title would be uniquely identified and could be explicitly linked to by an annotation overlay like the one described here.

<sup>136</sup> If the analytical annotation is removed, the textual coordinate system can be preserved by replacing the analytical word-level elements (`<w>`, `<num>`, `<pc>` and `<c>`) with semantically empty `<seg>` elements which inherit their `@xml:id` values.



to particular local difficulties and special-case situations [...], the intelligent short-cut” is not allowed to “countermand the nonintelligent, as well as counterintuitive, procedural logic of the electronic medium” (342). As a check for the preservation of machine-readability and capability for automatic processing, all of the editorial presentations of edited content included in this thesis—the diplomatic transcriptions, parallel reading edition and browsable HTML edition included on the accompanying CD-ROM (see appendices B, C and D) are produced procedurally from the base data files and annotation overlays with no manual intervention or post-editing, using algorithms that do not contain information or assumptions about specific phenomena in the source data but are designed purely on the basis of the annotation guidelines described in this chapter.<sup>137</sup>

This adherence to content-agnostic formal annotation mechanisms also makes the practices documented in this chapter highly generalisable, allowing for the accommodation of different text types simply through the definition of new attribute values (and perhaps the addition of some new elements for phenomena not encountered in the texts edited here) without needing to alter the annotation practices for the basic features shared by all historical manuscript documents. As Hockey (2004: 361) has argued, the “expense of creating electronic information” makes this kind of generalisability and the ability to satisfy many purposes a highly desirable trait for any digital editing framework, and has the added benefit of increasing the likelihood of its wider adoption. However, it should also be kept in mind that too wide and unrealistic claims regarding the general applicability of standards or editorial frameworks can easily create distrust in the validity of annotations standards in general, and lead to the situation observed by Lehmberg and Wörner (2008) among corpus compilers, where “many researchers still do not bother with standards or do not see their importance or the benefits of using them” (484). For this reason, the present framework is not claimed to be appropriate for all types of digital diplomatic editions, but rather to a specific subset of digital editions prepared of historical documents for the purposes of corpus linguistics.

---

<sup>137</sup> This means that the scripts used to generate the editorial output—included in appendix E—contain instructions for processing a variety of specific annotation variants that do not occur in the present edition but are allowed by the specifications described in this chapter.



**Part IV**

**Analysis**



## Chapter 12

# Dialectal characteristics

The developing nature of the English language in the absence of a standard meant that scribes were usually more than simple copyists. Texts could be written in various dialects, and they could be rewritten to make them suitable for other dialects or for later stages of the language [...]. So the question of establishing the dialectal origin of the text is important. (Blake 1998: 66)

As was observed in section 6.2, the 15<sup>th</sup> century was characterised by rapid and fundamental changes in the linguistic features of English texts (Lucas 1998: 178), and this “developing nature of English language in the absence of a standard meant that scribes were usually more than simple copyists”, frequently rewriting texts “to make them suitable to other dialects or for later stages of the language” (Blake 1998: 66, see also subsection 2.3.3). This means that even if the dialectal origin of a text cannot be established with any level of certainty—as is often the case with 15<sup>th</sup>-century texts—the documentation and analysis of the scribe’s linguistic usage is nevertheless “particularly important” for texts of this period, as it can help us in “tracing the spread of standardization in written English” prior to the establishment of Early Modern Standard English (Lucas 1998: 171). For this reason, this chapter will also include a dialectal analysis of the six manuscript texts edited here, based on the *Linguistic Atlas of Late Middle English (LALME)*.

The potentially fragmented textual transmission of *discourse colonies* like recipe collections means that colonies copied from multiple sources are likely to represent not only the multiple layers of relict spellings typical to non-authorial manuscript copies (see e.g. Lucas 1998: 171-2) but also several different transmission histories for different parts of the colony and are thus unlikely to make up a consistent linguistic profile. This means that the ideal unit of analysis from a purely theoretical point of view would be the individual recipe, being the largest unit that can be assumed to have been transmitted as an unchanged entity. However, the limited length and highly specific vocabulary of recipes means that establishing a sufficient linguistic profile for an individual recipe is problematic. For this reason, the analysis of the aggregate linguistic features of the whole collection is

here considered an acceptable compromise and a useful starting point for a more fine-grained future analysis.

Apart from the specific problems associated with the discourse colony, it is also often difficult or impossible to tell whether a given dialectal feature found in the text is the result of the scribe “translating” the text into his own dialect, a stray relict from an earlier exemplar, or something in between (Benskin and Laing 1981: 56). This means that the dialectal profile of the text—even if seemingly consistent—may equally well represent the area of origin of the ‘original author’ or any of the text’s subsequent copyists, or none of these, as its internal consistency can result either from the literal copying of an earlier—consistent—exemplar or from the consistent translation of the exemplar into the scribe’s own language (Mills 1998: 187), although McIntosh (1989a) considers the latter to have been much more common in the late 14<sup>th</sup> and early 15<sup>th</sup> centuries.<sup>1</sup> Unfortunately, as Mills (1998) points out, the period was equally characterised by a scribal practice somewhere between these extremes, with scribes neither consistently following the forms of their exemplar nor consistently replacing them with their own, leading to ‘relicts’ of the exemplar’s language being present in the copy “in a random fashion” (198). However, whatever the origins of the dialectal forms exemplified by the manuscript features, they can be taken to represent the passive linguistic repertoire of their immediate animator, or an acceptable use of language in the discourse community in which the manuscript copy was produced. For these reasons, the present initial analysis of the dialectal characteristics of the six *Potage Dyvers* versions is based on the aggregate analysis of the entire collections, supplemented by a brief examination of the occurrence patterns of minority forms, examining the degree to which the minority forms of different linguistic items coincide in the same recipes, possibly indicating a textual histories divergent from the majority of the collection.

## 12.1 Methods of analysis

The raw data for the linguistic profile of each *Potage Dyvers* version was collected by using an XSLT transformation (“Extracting\_LALME\_forms.xsl”, included in appendix E) which counted the occurrences of the different forms attested for the items included in the questionnaire used for the *LALME* on the basis of the regularised spelling and word-class information annotated for each word (see subsections 10.3.1 and 11.9.1).<sup>2</sup> For items requiring contextual differentiation, also a KeyWord-In-Context (KWIC) concordance of the instances was produced in order to allow the manual association of individual instances with the correct *LALME* questionnaire sub-item.<sup>3</sup> For creating the linguistic profiles, only forms attested in

<sup>1</sup> Even the linguistic profile of a single scribe can contain idiosyncratic or ‘mixed’ features as a result of his or her personal history. For example Mills (1998: 197) has observed that the language of John Gower is “a mixture of forms characteristic of southwest Suffolk and Northwest Kent”, correlating with the land holdings of his family, revealed by historical evidence.

<sup>2</sup> Unfortunately, not all of the items in the *LALME* questionnaire could be reliably located using the information available in the edition, which means that only 475 of the total of 818 items and sub-items of the *LALME* questionnaire were extracted for the present analysis.

<sup>3</sup> The manually edited raw data for the linguistic profile of each manuscript version, used for the analysis, is included in the folder Quantified LALME profiles in appendix G in the form of XML

the hand of the original scribe were included in order for the profile to reflect the usage of a single scribe, following the policy of McIntosh, Samuels and Benskin (*LALME*: vol.3: ix). The complete linguistic profiles for each of the six manuscript versions of *Potage Dyvers* are presented in appendix G on the included CD-ROM.

In order to determine the most likely area of origin for each text, the automated “Fitting” tool included in the electronic version of the *LALME* (Benskin, Laing et al. 2013) was initially used to analyse the dialectal profiles produced for the texts. According to the rather sparse documentation, the tool implements the ‘fit’ method described by Benskin (1991) to compare the forms attested in the new text to those of the texts localised in *LALME* to show which of the linguistic profiles included in *LALME* are closest to the new text and thus represents the most likely localisation for the text. However, since the tool does not fully implement Benskin’s original geographically-based ‘fit method’ but operates simply on the level of individual linguistic profiles (LPs), it does not take into account attestations of features in closely neighbouring profiles when evaluating the likelihood of a region as the origin of the analysed text.<sup>4</sup> This means that closely spaced clusters of LPs, each of which share different forms with the analysed text are not indicated as likely areas of origin, even if they together exhibit matches for all of the items analysed.

These problems, together with the fact that the *eLALME* Fitting tool is essentially a ‘black box’ that does not provide any way of following the process of elimination or of evaluating the results, made it unsuitable for the purposes of localising the *PD* manuscripts and necessitated the development of an alternative semi-automated implementation of the ‘fit’-technique. The fundamental problem of the fit-technique underlying the above-mentioned issues of the automatic tool is its fundamentally binary nature with regard to the individual forms, as it is based on the classification of LPs to those which exhibit the particular form (or group of forms) and those which do not. This creates a problem for those items which are not attested at all, since we have no way of knowing whether the text would employ the same form as the text being analysed or a different one but must nevertheless either include or exclude it in the set of potential matches. While the ‘inclusive’ approach used by the *eLALME* tool leads to the inclusion of spurious matches,<sup>5</sup> the converse approach of excluding all LPs which do not contain the item in question—the ‘exclusive’ approach (adopted here with some modifications)—can in the extreme case lead to any LPs that do not exhibit all of the forms included in the analysis being excluded as potential matches even if they are identical in terms of all items shared with the text being analysed. While the side effect of the ‘inclusive’ option might at first glance seem less problematic—after all, it might seem safer to avoid excluding an area for which there is no evidence

---

files produced by the above mentioned XSLT transformation and manually edited to disambiguate ambiguous forms, to remove false hits, and to consolidate lists of individual words exhibiting specific affixes into lists of those affixes.

<sup>4</sup> Instead, it simply compares each LP separately to the defined profile, decreasing the ‘similarity level’ of an LP only when a selected questionnaire item occurs with a form different from that in the profile being compared

<sup>5</sup> In this approach, the non-attestation of an item in a LP is essentially equated with the attestation of a matching form for the item, which in turn means that profiles which do not contain forms for any of the questionnaire items selected for the analysis register as maximally similar to the analysed text, even though they have nothing in common with it. Considering the number of *LALME* LPs for which only a few items are attested, this problem is not a marginal one.

either way—this is in fact a fallacy. As McIntosh, Samuels and Benskin (*LALME*: vol. 1, (check pages of chapter 3)) observe and the linguistic profiles of the *Potage Dyvers* also demonstrate, a scribal repertoire may well include several forms for a single item, and the presence of one form for an item in a LP cannot automatically be taken as evidence for the absence of another form any more than the complete lack of forms for that item. Thus, if we adopt the inclusive approach, we really have no criteria for excluding any LPs or areas unless we have strong prior evidence that specific forms are mutually exclusive and very unlikely to occur together in any area. Unfortunately, this would leave us very few forms to base our diagnoses on, since most of the forms in *LALME* have areas of attestation that overlap at least to some degree, and one form thus cannot be used to exclude the other. For this reason, the methodology used here makes no distinction between the absence of the form being analysed and the absence of *all* forms for the item, taking the positive presence of the analysed form as the only form of substantial evidence available to us.

In order to mitigate the effects of this essentially exclusive approach and to decrease the risk of excluding viable areas of origin on the basis of non-attested items, several precautions are observed. First of all, the present methodology—unlike the automatic fitting tool of *eLALME*—does not restrict the influence of a matching form to the individual LP but extends it to the wider area surrounding the likely geographical location of the LP. This means that matching forms for different items that are attested in closely situated LPs act to cumulatively increase the likelihood of the entire area as the dialectal source of the text, and that the occurrence of the form in a single LP in an area is enough to prevent the exclusion of the whole local area. Second, only items with reasonably widespread attestation in both *LALME* and the analysed text are used as the basis of localisation. This is intended to prevent the unwarranted exclusion of areas on the basis of forms that occur in the text as the result of scribal error or *Mischsprache* or forms that do not occur in any form in the majority of LPs and thus have little diagnostic value. Simultaneously this limitation also shifts the focus of the analysis towards ‘major’ linguistic forms which are likely to reflect the conscious language use of the scribe, rather than minor—and potentially incidental—ones.

In terms of practical implementation, the method used here replaces the physical overlays superimposed on the printed version *LALME*—recommended by Benskin (1991)—by *exclusion templates* based on item maps generated by the “User-Defined Maps” function of *eLALME* for those forms (or groups of forms) which are well-attested in both the text and the *LALME* maps, and have an analytically useful geographic distribution (Benskin 1991: 17).<sup>6</sup> The item maps were then converted into two kinds of digital overlays: a *dot map* (see *Figure 12.1a*) with a black dot representing each LP in which the form in question occurs, and an *exclusion template* (see *Figure 12.1b*) which covers the areas *not* containing matching LPs

<sup>6</sup> In terms of the number of occurrences, only forms (or groups of similar forms) that occur at least ten times in the text and in at least ten LPs in *LALME* were used in order to ensure the robustness of the analysis. In terms of geographical distinctiveness, only forms whose geographical distribution excludes at least some of the total area in which the item represented by the form is attested are used for the analysis. It should be kept in mind, however, that since the *eLALME* visualisation tool does not take the relative frequencies of items into account, the maps do not distinguish between LPs exhibiting the form as a minor form from ones where it is a major form.



with a black overlay, leaving the areas surrounding matching LPs transparent,<sup>7</sup> thus essentially implementing a more fine-grained version of the isoglosses used by Benskin (1991). While the relatively large radius used for the area considered to be represented by an individual LP serves to mitigate the effects of items not appearing in individual LPs through chance or sampling practices and thus to prevent spurious exclusions, this approach does not take into account the *density* of attestations, considering all areas where the form is found to be equally likely matches.<sup>8</sup> Consequently, the only criterion used for judging the likelihood of an area as the potential origin of a manuscript is the fact that a given dialectal form used in the MS has been found in use within it, the most likely areas of origin thus being those where all of the forms frequently used by the analysed text have been found to occur.

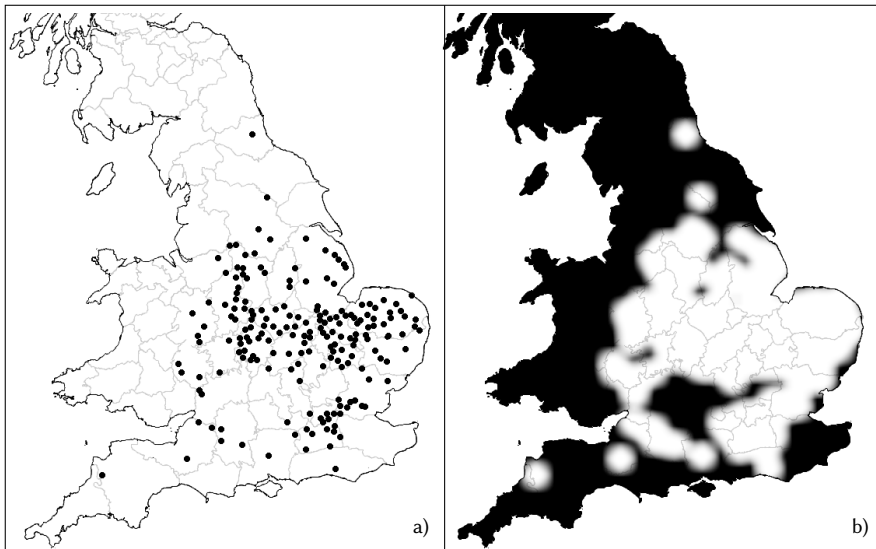


Figure 12.1: Examples of a dot map (a) and an exclusion template (b) for a single LALME questionnaire form.

## 12.2 Dialectal features of the manuscript versions

In general, the dialectal features of all the manuscript versions of *Potage Dyvers* would seem to point at a more southerly than northerly origin, with southern features like the past participle prefix ⟨y-⟩/⟨i-⟩ and initial ⟨h-⟩ for ⟨wh-⟩ being prevalent. This would also seem to be supported by perhaps the oldest and best

<sup>7</sup> Technically, this has been achieved by creating a mask from the dot map, extending, smoothing and feathering it and then making the area covered by the mask transparent. This process and its results are illustrated by the digital images included in appendix G, representing the geographical distribution of each individual dialectal form used for the cumulative maps in this section.

<sup>8</sup> While areas where a form occurs in the majority of LPs are more likely to produce texts with that form, the converse, namely that an individual text would be more likely to originate in that area cannot be inferred from this, since the text could equally well be a rare example as a common one.

known dialectal distinction of the English language, namely the almost exclusive use of the form *eyryn* for ‘eggs’ in all versions but MS As, traditionally considered a characteristically southern or Kentish form in Middle English, if only on the basis of an anecdote provided by Caxton in the preface to his *Eneydos*.<sup>9</sup> The following subsections will describe the dialectal features of each manuscript version in detail, based on the analysis methods described above.

### 12.2.1 British Library MS Harley 4016

In terms of the orthographic realisations of individual words, the linguistic profile of MS Harley 4016 (abbreviated as MS H4016) contains less variation than the other versions, the average number of forms per questionnaire item attested in the text being less than two.<sup>10</sup> For items which exhibit more than one form, the dominant form accounts for more than four fifths of all occurrences on average.<sup>11</sup> Based on a cursory observation of the distribution patterns of the forms listed in *Table 12.1* the language of this manuscript version would seem to represent either a Midlands or Essex variety. The north as an area of origin is precluded by forms like *togidre/to gidre* for ‘together’ (item 268), *porgh* for ‘through’ (item 55), *pen* for ‘then’ (item 30), *wilt* for the second person singular of ‘will’ (item 24-20), and *hit* for ‘it’ (item 8). While many of the forms would allow for an East Anglian or even a Southern origin, the form *litul/litull* (item 191) is characteristic of the West Midlands and does not occur in western East Anglia (which would otherwise be a plausible candidate area), and the form *to gidre* (item 268) does not occur in the South, apart for a single occurrence in the Southwest.

While the areas of most frequent attestation for these items are located in various places across the Midlands, the application of the fit technique to the forms listed in *Table 12.1* results in three areas where all of the analysed forms co-occur in close proximity, making the linguistic profile of MS H4016 the most dialectally indeterminate of all the *PD* versions. As can be seen in *Figure 12.2*<sup>12</sup>, the first of these areas is located in southern Herefordshire near the Welsh border,<sup>13</sup> the second in southern Northamptonshire and the third in London. *Figure 12.3* provides two alternative visualisations of the same data, one of them representing the degree to which individual LPs match the dialectal features of MS H4016, and the other representing a graduated and more detailed version of the ‘fit’ technique

<sup>9</sup> The anecdote is reproduced for example in Baugh and Cable (2002: 195-6).

<sup>10</sup> It should be noted that these figures on the orthographic variation of dialectally relevant items are based only on the types included in the *LALME* profiles compiled for the texts, and not on their entire lexis.

<sup>11</sup> The average ratio between the frequency of the dominant form and the combined frequencies of the rest of the forms is roughly 4.4, varying between 0.4 and 38 for individual items. All of the attested forms along with their absolute quantities are listed in the file “MSHarley4016\_LALME\_profile\_-quantified.xml” included in appendix G.

<sup>12</sup> This map, along with those for the other five manuscript versions (*Figures 12.4, 12.7, 12.9, 12.11, and Figure 12.13*) has been produced by superimposing the exclusion templates for all the items listed in *Table 12.3* (and the corresponding table for each MS version), leaving visible only the area of the underlying map which falls within a certain radius of an LP exhibiting each of the listed forms. Of the forms occurring more than ten times in this text, the form *togidur* for ‘together’ (item 268) was excluded from the analysis as it only occurs twice in *LALME* (in Warwickshire and Buckinghamshire) and was not considered to provide a reliable basis for localisation.

<sup>13</sup> This is in fact very close to the area indicated as the best fit for the MS H279 version (see subsection 12.2.2) below.

No. ITEM form (#)
6 IT hit (519)
8 THEM hem (466)
24-20 WILL (2sg) wilt (17)
30 THEN then (196)
30 THEN þen (140)
55 THROUGH thorgh (84)
55 THROUGH þorgh (25)
56 WHEN whan (76)
100 BUT but (36)
137 FIRE fire (65)
140 FLESH flesshe (20)
155 GOOD gode (14)
155 GOOD good (33)
155 GOOD goode (25)
160-20 HAVE (inf) haue (13)
191 LITTLE litel/litell (10)
191 LITTLE litul/litull (83)
268 TOGETHER to gidre (12)
268 TOGETHER togidre (80)
268 TOGETHER togidur (10)
67 (S) PPL PREFIX y/y- (34)
115-20 (S) DO (2sg) doest (14)
172 (S) HIS his (51)
277 (S) UNTIL til/till (68)



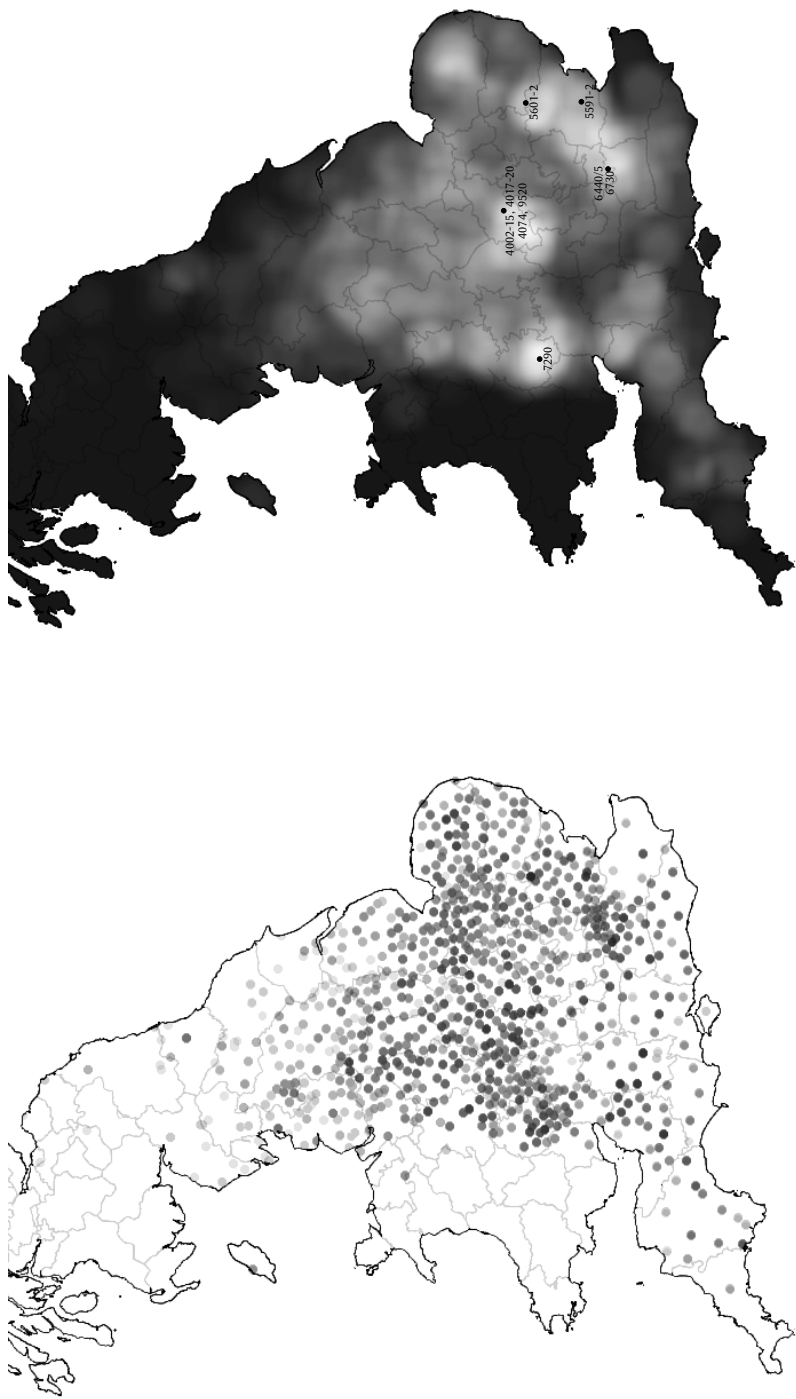
**Table 12.1:** *The forms used for analysing the dialectal ‘fit’ of MS Harley 4016.*

**Figure 12.2:** *The area of occurrence shared by all of the features listed in Table 12.1, marking the most likely area of origin for MS Harley 4016.*

with a lighter colour indicating greater agreement with the linguistic profile of the analysed text.<sup>14</sup> As the latter map reveals, also the general area of eastern Essex—which will be found to match many of the other versions of *PD*—shares many of the forms witnessed here. The graduated exclusion map also highlights those LPs in the likely areas of origin that display the closest affinity to the profile of this text.

In the Herefordshire region, the individual LP providing the closest match is LP 7290, derived from the BL MS Additional 4698 version of the herbal *Agnus Castus*. The strength of the fit, however, is not solely due to this single text, but rather due to the large number of other nearby profiles that also exhibit a significant number of the diagnostic forms used in the analysis. The area of match coinciding with London is slightly weaker, the LPs exhibiting the various forms being slightly

<sup>14</sup> For the composite dot map, the dot maps for the individual forms were simply reduced in opacity and superimposed on top of each other, resulting in the cumulative darkening of the LPs depending on the number of forms they have in common with the analysed text (the opacity of individual maps has been adjusted according to their total number, resulting in a cumulative opacity of c. 90% for a fully matching LP). The graduated exclusion template has been produced using a similar method, using more detailed exclusion templates (with a smaller radius around each matching LP) than for the absolute version, adjusting their opacity in the same way as for the dot maps, and superimposing them. The combined dot maps and detailed exclusion templates for each individual feature—for both this and other versions of the *PD*—are also included in appendix G.



**Figure 12.3:** A graduated dot map showing the degree of similarity between MS Harley 4016 and the localised LALME profiles (darker dots representing more similar LPs), and a graduated sum of exclusion templates (darker colour indicating decreased similarity), marked with the most similar individual LPs located in the areas of greatest overall similarity.

farther apart, as seen by the slightly grayed out appearance of the white spot in *Figure 12.2*, and there are no closely matching individual LPs in the immediate vicinity of London.<sup>15</sup> In the Northamptonshire area, the closest individual match is provided by the series of Northampton borough records making up LPs 4002–9, 4011–5, 4017–20 and 9520, along with the play manuscript represented by LP 4074, all localised to Northampton, while the closest matches in Essex are the two LPs (5601–2 and 5591–2) that also match many of the other *PD* manuscripts.

Although shown as the largest area of overlap in *Figure 12.2*, the status of the Northampton region is undermined by the fact that the southern past participle prefix, frequently attested in the text, is only recorded as ‘present’ (value X) for this area and we cannot be certain of its form.<sup>16</sup> The Essex area on the other hand would seem to be excluded by the characteristically West Midlands form *litul/litull* for ‘little’ (item 191), which does not occur there.<sup>17</sup>

In terms of individual dialectal features, the three areas indicated by *Figure 12.2* are minimally defined by only three items, the rest of the forms listed in *Table 12.1* merely strengthening the localisation and slightly narrowing down the three areas. The forms *to gidre* and *togidre* for ‘together’ (item 268) exclude not only the North but also much of the South and the northern Midlands, occurring mostly in East Anglia, Essex and the southern Midlands. The form *litul/litull* for ‘little’ (item 191), on the other hand, excludes East Anglia and Essex, occurring almost exclusively in the West and Central Midlands. Finally, The form *wilt* for the second person singular of ‘will’ (item 24-20) eliminates the southern tip of the East Midlands on the shores of the Wash and much of the Central Midlands.

Unfortunately even an examination of the less frequent forms occurring in the text does not provide conclusive evidence for deciding between the three possible areas of origin, as the majority of them are dialectally neutral and occur more or less equally in or near all of the three areas.<sup>18</sup> Since the overall impression given by the language of this text is quite ‘modern’, the forms used largely corresponding to their modern counterparts, the indeterminacy of the dialect of this version is most likely an indication of the standardised and ‘neutral’ nature of the language, also supported by the relative lack of variation mentioned above. A high degree of standardisation with generic Midlands forms could in turn be taken as an indication that that text was produced in London, perhaps by a scribe used to the standardised language of official documents.

<sup>15</sup> The nearest close matches are LPs 6440/5 and 6730 to the west in southwestern Middlesex and LP 5591–2 to the east in southern Essex.

<sup>16</sup> The isogloss of the recorded *y* value runs much farther south, excluding the Northamptonshire area but including London and southern Herefordshire.

<sup>17</sup> It is, of course, possible that these features are either contaminations from the language of a scribe with a Southwestern background or relicts from a Southwestern exemplar, although the fact that *litul/litull* is by far the dominant form for ‘little’ in the text makes this rather unlikely.

<sup>18</sup> In addition to the forms included in the above analysis, the only infrequent forms that seem to differentiate between the three areas to any degree are *togidur* for ‘together’ (item 268) and *ren* for the infinitive of ‘run’ (item 233-10), both of which have only a few attestations in *LALME*, the former in Central Midlands (Warwickshire and Buckinghamshire) and the latter in Essex and in Somerset. However, their paucity of attestations makes it difficult to draw any conclusions from them.

12.2.2 British Library MS Harley 279

In terms of the average level of orthographic variation, the linguistic profile of MS Harley 279 (abbreviated as MS H279) is more varied than that of MS H4016. Similarly to most of the other versions (see below), the average number of forms found per *LALME* questionnaire item is slightly under two and a half. While the ratio of the dominant form to minor forms varies, just as in the other versions, the dominant forms in this text are on average ‘stronger’ than in any other version, accounting for a greater proportion of instances.<sup>19</sup> The general impression based on the geographic distributions of the questionnaire items listed in *Table 12.2*, is that of a Southern language variant, as there is quite a number of forms that occur predominantly in the south, such as *þey* for ‘they’ (item 7), *hem* for ‘them’ (item 8), *moche* for ‘much’ (item 16), *schal* for ‘shall’ (item 22), *nat* for ‘not’ (item 36), *aʒen* for ‘again’ (item 37), *lytel* for ‘little’ (item 191), and *togedere* for ‘together’ (item 268).

No. ITEM form (#)
7 THEY þey (55)
8 THEM hem (639)
16 MUCH moche (11)
24-20 WILL (2sg) wolt (22)
30 THEN þan (394)
30 THEN þen (169)
33 IF ʒif (48)
41 WHILE whyle (31)
46 NOT nowt (28)
55 THROUGH þorw/þorwe (97)
56 WHEN whan (77)
100 BUT but (38)
137 FIRE fyre (71)
140 FLESH fleys- (42)
155 GOOD gode (122)
160 HAVE haue (15)
191 LITTLE litel (14)
191 LITTLE lytel (33)
191 LITTLE lytil (53)
268 TOGETHER togedere (13)
268 TOGETHER togederys (107)
275 TWO to (13)
67 (S) PPL PREFIX y- (91)
277 (S) UNTIL tyl (49)
277 (S) UNTIL tylle (21)

**Table 12.2:** The forms used for analysing the dialectal ‘fit’ of MS Harley 279.



**Figure 12.4:** The area of occurrence shared by all of the features listed in Table 12.2, apart from the form *nowt* for ‘not’ (item 46), marking the most likely localisation for MS Harley 279.

While MS H279 does not exhibit an unusual amount of variation within an

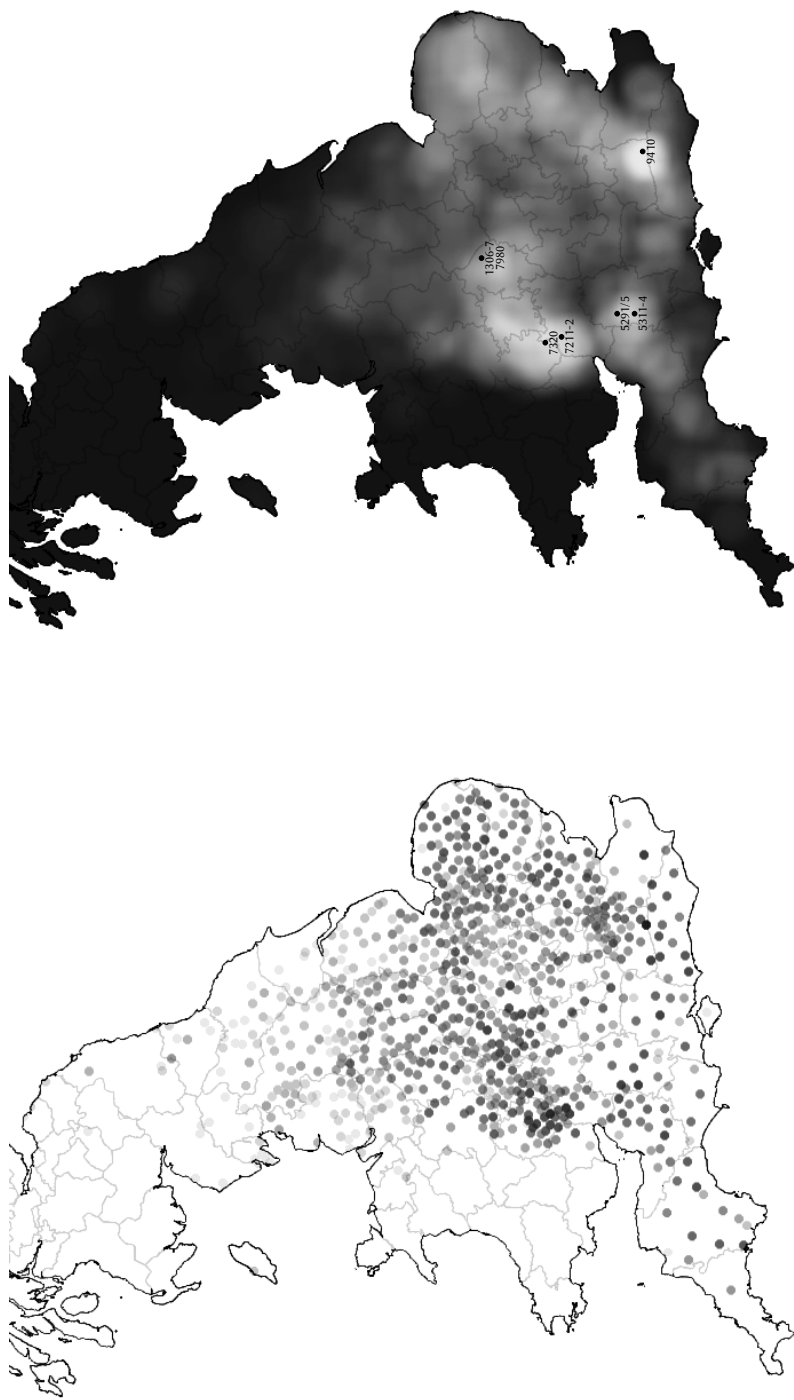
<sup>19</sup> The ratio of the dominant form to all other forms is 5.7 on average, varying between one third and 54. All of the attested forms along with their absolute quantities are listed in the file “MSHarley279\_LALME\_profile\_quantified.xml” included in appendix G.

individual item, the linguistic profile of the text would seem to be more heterogeneous than the others in terms of the dialectal areas represented by the forms used for different items, making the process of ‘fitting’ the text an unusually difficult one. This is slightly unexpected, since MS H279 is a special case among the *Potage Dyvers* manuscripts in that it has actually been localised in *LALME* (based on Austin’s 1888 edition) and has its own language profile, LP 9410, which would be expected to provide an exact match in the analysis. However, applying the fit technique to the forms listed in *Table 12.2* by superimposing their exclusion templates results in total exclusion, indicating that there is no region in which all of the forms would occur together.<sup>20</sup>

If we look at the alternative visualisations presented in *Figure 12.5*, which show LP 9410 as a very close—although not a complete—match, we can see that even with LP 9410 itself included in the analysis, the area surrounding it does not in fact provide the closest match in terms of the forms analysed. If we eliminate its influence from the results—as should properly be done when trying to place the text—we can see (*Figure 12.6*) that the area in which it has been localised in *LALME* does not provide a notable match to the forms used in this text and is actually excluded by the isoglosses of several of the forms listed in *Table 12.2*, including *wolt* (item 24-20), the forms for ‘flesh’ (item 140) beginning with *fleys-*, the present form *haue* (item 160)<sup>21</sup>, and the form *lytil* for ‘little’ (item 191). While none of these exclusions are particularly strong—all of the forms are attested in nearby regions—the area centred on the Herefordshire–Gloucestershire–Worcestershire

<sup>20</sup> The elimination of the area of the LP assigned to the manuscript in *LALME* from the exclusion map is explained by two factors. First of all, an examination of the *LALME* profile and the profile used here (see appendix G) reveals that there are several forms and even entire questionnaire items that did not occur in the sample used to construct LP 9410 but do nevertheless occur in the manuscript, most notably the (southern) past participle prefix (item 67) and the present forms of ‘have’ (item 160). An examination of the geographic distribution of these forms using the “User-defined Maps” function of *eLALME* reveals that despite its ubiquitous nature, the present forms of *have* are in fact been recorded for extremely few southern texts, although it is very frequently recorded in northern LPs and is unlikely to have been that much rarer of a verb in southern ones. The past participle prefix *i* or *y*, on the other hand seems to be recorded somewhat inconsistently, being sometimes recorded as the actual prefix (e.g. *y*), sometimes just as being present (value X), and sometimes not at all (as in LP 9410). This means that in the implementation used here, in which only the positive presence of a form spares the LP from elimination, there is a constant danger of spuriously eliminating areas where the form could well be current but is not recorded in the LPs for one reason or another. However, as mentioned above, the contrary solution of eliminating only those LPs and areas which exhibit an alternative form leads to the opposite problem, namely that LPs which have few or no items in common with the analysed text are shown as close matches. This problem, which could be called ‘inconclusivity of negative results’ is not specific to *LALME*, but rather common to research data based on a samples representing larger populations: the absence of a feature in a sample cannot be taken as evidence that the feature does not occur in the population, as this would be a form of *argumentum ad ignorantiam*. While there is no obvious solution to this particular problem, the inclusion of wider areas surrounding a matching LP instead of just the individual LP, as well as the use of only well-attested items and forms serve to mitigate the problem by allowing neighbouring LPs to compensate for the absence of a feature in a specific LP in defining the areas of best fit. This approach is based on the notion that if a form—even for a rarer item—is current in an area, it is likely to occur in at least *some* of the texts from that area, increasing the odds that the area is registered as a potential match even if only one in a number of LPs exhibits the form in question. The second factor resulting in the exclusion of the area around LP 9410 is an illustration of the conditions in which these precautions fail: the area is excluded in this analysis simply because there are not that many other LPs around LP 9410 that match the profile of the analysed text to any significant degree, and would thus preserve the area from elimination.

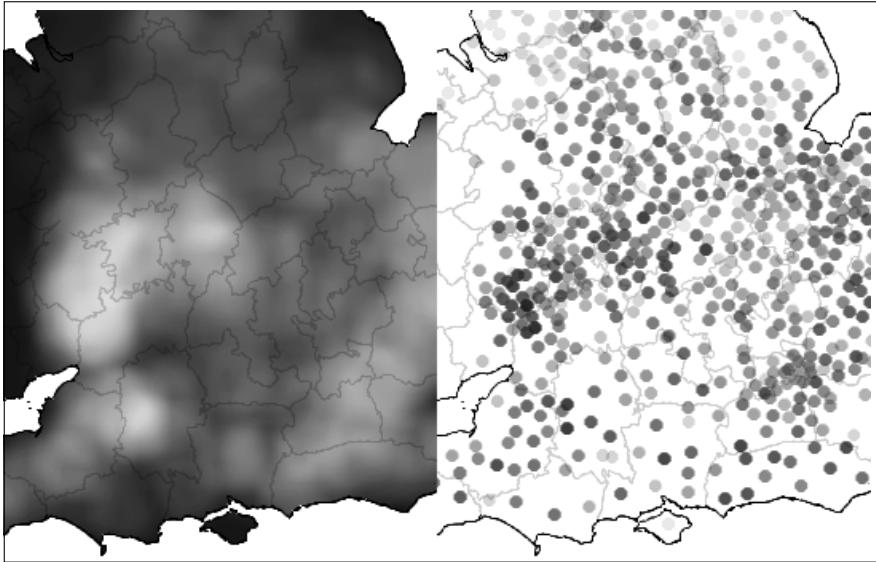
<sup>21</sup> Although this might be merely an anomaly of the compilation practice of the southern profiles.



**Figure 12.5:** A graduated dot map showing the degree of similarity between MS Harley 279 and the localised LALME profiles (including the profile assigned to it on the basis of Austin (1888)), and a graduated sum of exclusion templates, marked with the most similar individual LPs located in the areas of greatest overall similarity.



border would nevertheless seem to offer a significantly better match for the text, being excluded only by the form *nowt* for ‘not’<sup>22</sup> (item 46), a distinctly eastern form occurring almost exclusively in northern East Anglia and in the area around London<sup>23</sup>, although it does have isolated occurrences all around the country, most likely representing East Anglian influence. If we apply the fit technique to all of the forms listed in *Table 12.2* except for *nowt*, we can see that all of the other features analysed do in fact occur in this region, as shown in *Figure 12.4*.



**Figure 12.6:** A detail view of the dot map and graduated exclusion map of MS Harley 279 with LP 9410 excluded.

Since there are also some other forms (besides *nowt*) that occur most frequently in East Anglia, such as the forms for ‘flesh’ (item 140) beginning with *fleys-* and the form *lytil* for ‘little’ (item 191), it is also prudent to examine the fit of the text to this region. The forms that would seem to exclude East Anglia—or more specifically the Suffolk–Norfolk border—as a likely area of origin are *togederys* for ‘together’ (item 268) and the *tylle* for ‘until’ (item 277) in the southern questionnaire. While *nowt* is the dominant form used for ‘not’ in this version, *togederys* is used even more exclusively and much more frequently, and while *tylle* is a minority form, it occurs almost as frequently as *nowt*. The rest of the forms analysed here can be considered inconclusive in terms of the West Midlands–East Anglia dimension, as they occur in both, although with varying frequencies.

As McIntosh, Samuels and Benskin (LALME: 12-3) point out, the occurrence of these kinds of anomalous forms is most likely an indication of some form of dialectal mixture. Considering that the non-conformant form originates from an entirely different part of the country, its occurrence is most likely due to the second type of mixture outlined by McIntosh, Samuels and Benskin (LALME: 13),

<sup>22</sup> In the present edition, this form, along with its cognates are regularised into ‘nought’ (in the C.2 sense of the *OED*).

<sup>23</sup> See the individual item maps in appendix G.

i.e. resulting from several layers of copying, as the distance between the likely areas means that the language variety is unlikely to be a ‘border dialect’ or the idiosyncratic dialect of a single scribe, and the features which are exclusive to these two areas cannot be considered standard forms. The view of this particular form being a *relict* form from a hypothetical East Anglian exemplar (see section 12.3 below) is supported by the fact that the other forms of ‘not’ used in this text, *noʒt* and *not* (which together outnumber the *nowt* form) seem to be native to the Herefordshire–Gloucestershire–Worcestershire area. One explanation for the survival of such a relict form might be offered by the observation that the area indicated by Figure 12.4 also exhibits the forms *nowght* and *nowʒt*, which could have made the form *nowt* occurring in the exemplar familiar enough to a scribe from this area to be copied without needing to be systematically substituted with a more local variant.

The forms that minimally delimit the general area indicated by Figure 12.4 at the Hereford–Gloucestershire–Worcestershire border are *togederys* for ‘together’ (item 268), which seems to only occur in the very south, with one instance to the east of London and a cluster of occurrences in the north of Gloucestershire, *haue* for the present forms of ‘have’ (item 160), which is not recorded for a large proportion of south (although this may well be an anomaly, as pointed out above in footnote 20 on page 549), and the forms of ‘flesh’ (item 140) beginning with *fleys-*, which seem to be limited to East Anglia and the southern part of the Midlands, along with sporadic occurrences in the South. The closest match of the individual LPs in this area is provided by LP 7320 in the southeastern end of Herefordshire, which is surrounded by a large number of other LPs sharing many of the forms, the most notable of which is LP 7211/2 on the Gloucestershire side of the border. Further south, LPs 5291/5 and 5311/2/3/4 provide relatively close matches, while for example East Anglia provides no close matches.

Of the less frequent forms not included in the above analysis, the major forms for ‘her’, *hyre* and especially *hure* are very characteristic of the West Midlands, the latter occurring almost exclusively in this area and around the mouth of the Severn, apart from a handful of attestations in Essex and Kent. Also the rare minority form *meni* for ‘many’ (item 13), occurring once in the manuscript, is attested in four LPs in *LALME*, three of which are located in the west (Herefordshire, Gloucestershire, and Somerset) and one in Kent. The majority form used in the manuscript for the second person singular of ‘shall’, *schalt*, is attested mostly in LPs from the West Midlands, with a major concentration in southern Herefordshire, at the exact point indicated by Figure 12.4. Also the minority form of ‘might’ (item 54), *myʒth*, occurring twice in the manuscript is very frequent in southern Herefordshire and the surrounding area. Taken together, these minority forms also reinforce the placement of MS H279 in the western edge of the West Midlands, with some East Anglian influences inherited from an exemplar, rather than in southern Surrey, where it was placed in *LALME*.<sup>24</sup>

<sup>24</sup> While there is no external evidence about the geographical origin of the manuscript, a western origin could also be seen to offer an explanation for the inclusion of a bill of fare attributed to the wedding of the Earl of Devon, which would have been of more local interest in the western part of the country. This connection is, however, quite tenuous, as the wedding of an Earl could be considered to be of national interest, and the same bill of fare also appears in MS As (albeit without an attribution), which is here localised to either Essex or the central Midlands (see below).

12.2.3 British Library MS Additional 5467

The linguistic profile of the MS Additional 5467 (abbreviated as MS Ad) version of *Potage Dyvers* exhibits quite a wide selection of dialectal forms for many of the items on the *LALME* questionnaire, the average number of forms per item being more than three. In some cases one or two forms are clearly dominant, the rest occurring only a handful of times, while in others all of the forms occur with roughly equal—usually quite low—frequency.<sup>25</sup> Based on a general examination of the geographic distributions of the questionnaire items listed in *Table 12.3*, the text would seem to represent a Southern or Essex dialect. In terms of specific forms, the use of *wult* and *wolt* for the second person present singular of ‘will’, as well as the form *lite* for ‘little’ would seem to rule out the North as a likely area of origin. Furthermore, the limited occurrence of the form *flesshe* for ‘flesh’ in the South and Midlands, together with the use of the *y* prefix for past participles would seem to limit the likely areas of origin to either to the area around London, to Essex or to the southwest.

No. ITEM form (#)
8 THEM hem/heme (397)
24-20 WILL (2sg) wult/wolt (13)
30 THEN then/thene/thenne (354)
33 IF yf (33)
41 WHILE while/whille (25)
55 THROUGH thorow/thorowe (113)
100 BUT bot/bote (40)
137 FIRE fire (50)
137 FIRE fyre (11)
140 FLESH flesshe (16)
155 GOOD goode (57)
155 GOOD good (19)
191 LITTLE lite (32)
191 LITTLE litil/litill (52)
67 (S) PPL PREFIX y/y- (22)
115-20 (S) DO 2sg dost/doost (12)
277 (S) UNTIL til/till (31)
277 (S) UNTIL tyl/tyll (12)

**Table 12.3:** *The forms used for analysing the dialectal ‘fit’ of MS Additional 5467.*



**Figure 12.7:** *The area of occurrence shared by all of the features listed in Table 12.3, marking the most likely area of origin for MS Additional 5467.*

Applying the ‘fit’ technique by overlaying the exclusion templates for all the features listed in *Table 12.3* onto the *LALME* base map results in *Figure 12.7*, showing that the only areas where all of the forms occur in a nearby LP are the two areas covering western Essex, London and northern Surrey in the east and eastern

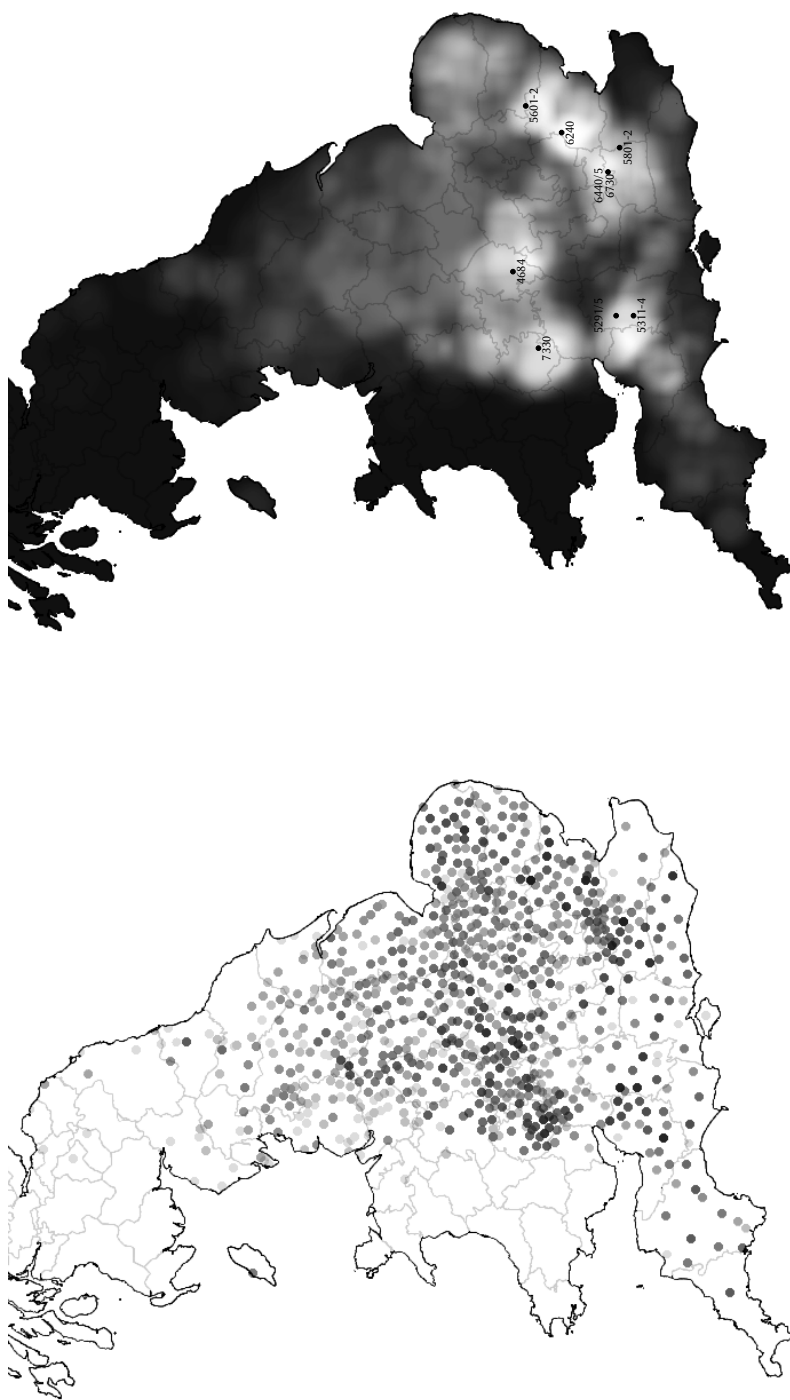
<sup>25</sup> All of the attested forms along with their absolute quantities are listed in the file “MSAdditional5467\_LALME\_profile\_quantified.xml” included in appendix G.

Somerset, western Gloucestershire and eastern Herefordshire in the west. Looking at the alternative visualisations in *Figure 12.8*, the LPs in the likely areas of origin that display the closest affinity to the profile of this text include LP 5601/2 and 6240 in western Essex, LPs 6440/5, 6730 and 5801/2 in the area south of London, LP 7330 in southern Herefordshire, and LPs 5291/5 and 5311/2/3/4 in western Wiltshire. If we analyse the contributions of all the listed forms to these maps, we can see that the minimal selection of forms that delimits these areas consists of the forms *thorow/thorowe* (item 55), *fire* (item 137), *flesshe* (item 140), *lite* (item 191), and the southern past participle prefix *y/y-* (item 67). Including the rest of the forms included in *Table 12.3* does not significantly restrict the area delimited by these forms, but confirms it as a possible area of origin for also those forms.

A general overview of the rare forms not included in the above analysis shows that most of the rare forms also conform to the general pattern, occurring both in Essex and East Anglia, and in the northern part of the Southwest. Unfortunately they do not provide sufficient evidence for deciding between the two likely areas of origin, painting a slightly inconsistent picture, although there would seem to be slightly more evidence for Essex than there is for Somerset. For example the form *thos* for ‘those’ (item 3) and the form *geder* for ‘gather’ (item 150-20) occur in east Essex (actually in LP 6240), but not in northern Somerset. Also the *-y-* forms for ‘run’ (*rynn*, *ryne*) seem to be limited to east, although this attribution is based on a single *LALME* occurrence of *rynn* in LP 5601/2 from the north border of Essex. On the other hand, the form *hy* recorded as a minority form for ‘they’ (item 7), is frequent in Somerset and Wiltshire but does not occur anywhere near Essex, and the minority form *ham* for ‘them’ (item 8) seems to be typical to West Midlands and Somerset, its only occurrence in Essex coming from the very south of the county.

Some of the more frequently attested items also exhibit minority forms that do not coincide with the two proposed locations, possibly indicating contamination from an exemplar (or from the native dialect of the scribe, depending on his method of copying). Examples of this include the rare form *tha* for ‘they’ (item 7) (which occurs only twice in *LALME*, in northern East Midlands), the unique occurrence of *lace* for ‘less’ attested only once in Cheshire, the majority form for the second person singular of ‘will’ (item 24-20), *wult*, which has very few occurrences in *LALME*, mostly in the West Midlands (with a single instance in Cornwall), and the forms for ‘about’ (item 70), namely *abowt* and *a bowt*, which occur mainly in the East Midlands, with two occurrences in Norfolk.

In terms of specific *LALME* profiles, the closest match in the area of western Essex is LP 6240 which is in fact very similar to the profile of this text. Although there are also other LPs that display an equal level of similarity, the fit of this particular LP is further strengthened by many of the forms also being attested in surrounding LPs, especially LP 6230, LP 6260, LP 6270 and LP 6530. In the west, the closest match is LP 5281-2, at the northeast corner of both Somerset and the area of best fit indicated by the analysis. Similarly to the area in east Essex, also this region is made a likely area of origin not so much because of a single closely matching LP but rather by a concentration of multiple profiles, each of them exhibiting some of the forms. LPs that are especially significant in establishing the fit include LP 5291/5, LP 5260, LP 5270, and LP 9420.



**Figure 12.8.** A graduated dot map showing the degree of similarity between MS Additional 5467 and the localised LALME profiles, and a graduated sum of exclusion templates, marked with the most similar individual LPs located in the areas of greatest overall similarity.

### 12.2.4 Bodleian Library MS Ashmole 1439

The dialectal forms making up the linguistic profile of MS Ashmole 1439 (abbreviated as MS As) are not quite as varied as those of MS Ad, the average number of forms found for every *LALME* questionnaire item being slightly under two and a half, resembling most of the other *PD* versions in this regard. Similarly to MS Ad, the internal relationships of the different forms vary, although on average, both the frequency difference between the dominant form and minor forms and the variation in this ratio is less pronounced than in MS Ad.<sup>26</sup> Based on a general examination of the geographic distributions of the questionnaire items listed in *Table 12.4*, the text seems to represent a variant associated with the county of Essex and the area surrounding London. In terms of specific diagnostic items, the forms *pey* and *pem* for ‘they’ (item 7) and ‘them’ (item 8), along with the forms *wol* and *wolle* for ‘will’ (item 20) exclude the North as a likely area of origin and label the language as either a Southern or Midlands one. While the limited occurrence of the second person singular of ‘will’ (item 24-20) in *LALME* makes it an uncertain diagnostic, the forms *wolt* and *wilt* would seem to indicate an origin either in the West or Central Midlands, the Southwest, or in the area extending from East Anglia to Sussex (although the form *panne* is more typical of the West Midlands). Also the forms for ‘through’ (item 55), namely *purgh/purghe* and especially *purw/purwe*—the latter of which occurs in a limited area, principally in East Anglia and Essex (with a few occurrences in the eastern West Midlands and in the Southwest)—would seem to indicate an Essex dialect, confirmed by the frequent occurrence of the prefix *y/y-* in past participles, the northernmost occurrences of which are found in Essex, north of London.<sup>27</sup>

Applying the fit technique to the frequently attested and geographically exclusive features listed in *Table 12.4* onto the *LALME* base map results in the cumulative exclusion template shown in *Figure 12.9*, revealing that the only area where all of the forms clearly co-occur in the same region lies in west Essex, extending from the border of Essex, Cambridgeshire and Suffolk to the southern border of the county east of London.<sup>28</sup> *Figure 12.10* provides two alternative visualisations of the same data, which show that the closest fit is provided by the area on the north side of the Thames estuary in southern Essex, east of London. While no single LP seems to be a perfect match, the strongest individual match is provided by LP 5591-2 in southern Essex, south of Chelmsford, on the eastern edge of the best-fit area indicated by the map. Other strong candidates at the northern and

<sup>26</sup> The average ratio between the frequency of the dominant form and the combined frequencies of the rest of the forms is only 3.2 (with a minimum of one third and maximum of 33) for MS As. All of the attested forms along with their absolute quantities are listed in the file “MSAshmole1439\_LALME\_profile\_quantified.xml” included in appendix G.

<sup>27</sup> The very limited occurrence of the prefix *y* may, however, be an artefact of *LALME* recording practices, as the presence of a past participle prefix is in many cases indicated simply by the positive marker *X*, which thus has a much wider distribution in the *LALME* maps. For this reason, the exclusion templates used for the analyses include also this form, even though it may also include other forms of the prefix.

<sup>28</sup> In addition to this area, all of the analysed forms also seem to marginally co-occur in northern Warwickshire, but since this area is only peripherally covered by the area of attestation for the frequently attested forms *wilt* and *wolt* (item 24-20), *thanne* (item 30), and *purw/purwe* (item 55), and *til/till* (item 277), as well as the southern past participle prefix *y*, it has been considered a less likely option than the area in western Essex.

No. ITEM form (#)
6 IT hit (808)
7 THEY þey (45)
8 THEM þem (370)
8 THEM them (172)
16 MUCH moche (11)
24 WILL wol/wolle (10)
24-20 WILL (2sg) wolt (14)
24-20 WILL (2sg) wilt (11)
30 THEN þanne (415)
30 THEN þan (67)
30 THEN þenne (61)
30 THEN thanne (21)
41 WHILE while (33)
52 THERE þer (11)
55 THROUGH þurgh/þurghe (73)
55 THROUGH þurw/þurwe (22)
56 WHEN whanne (69)
100 BUT but (40)
137 FIRE fire (44)
137 FIRE fyre (25)
140 FLESH flesshe (15)
155 GOOD gode (101)
155 GOOD good (22)
160 HAVE haue (11)
191 LITTLE litil (59)
191 LITTLE litel (54)
268 TOGETHER to gedre (105)
268 TOGETHER to gedir (26)
67 (S) PPL PREFIX y/y- (72)
277 (S) UNTIL til/tille (75)

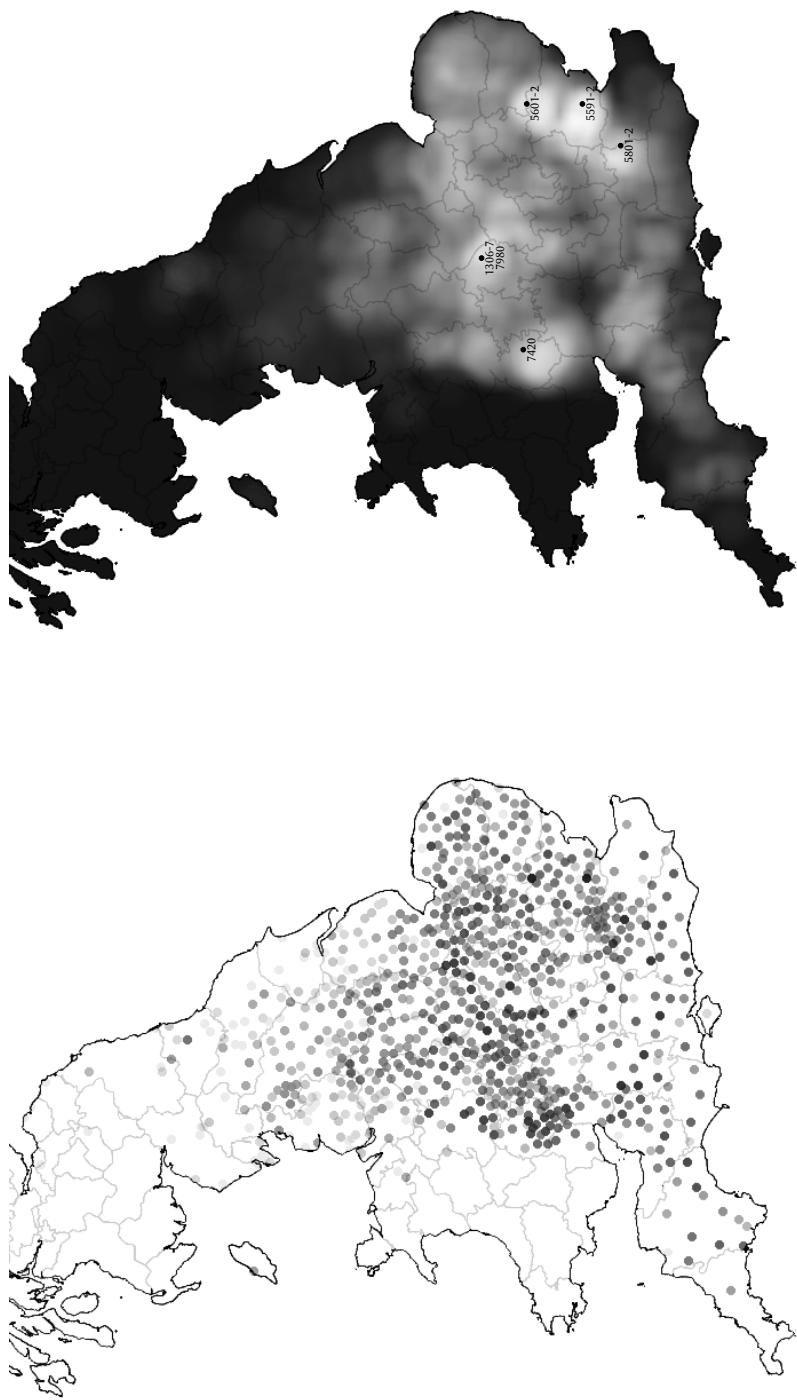


Table 12.4: The forms used for analysing the dialectal ‘fit’ of MS Ashmole 1439.

Figure 12.9: The area of occurrence shared by all of the features listed in Table 12.4, marking the most likely area of origin for MS Ashmole 1439.

southern ends of the likely area of origin are LP 5601–2 in northern Essex, near the borders of Cambridgeshire and Suffolk, and LP 5801–2 in eastern Surrey, south of London.

Analysing the contributions of all the forms listed in Table 12.4 to these maps reveals that the minimal selection of forms that delimits the general area of western Essex—albeit more loosely than the full list of forms—consists of the form *wilt* for the second person present of ‘will’ (item 24-20), the form *thanne* for ‘then’ (item 30), the prefix *y* for past participles (item 67), and most significantly, the form *þurw/þurwe* for ‘through’ (item 55). Adding the forms of all the remaining significant items does not significantly change the fit, but does narrow it down considerably. Among the forms listed in Table 12.4, the only form that does not occur in the area indicated by the other forms is the rare form *fleysshe*. Since the only occurrence of this form in *LALME* is in the LP of MS H279, described and contested above, this form has been considered to be insufficient grounds for



**Figure 12.10:** A graduated dot map showing the degree of similarity between MS Ashmole 1439 and the localised LALME profiles (darker dots representing more similar LPs), and a graduated sum of exclusion templates (darker colour indicating decreased similarity), marked with the most similar individual LPs located in the areas of greatest overall similarity.



abandoning the localisation indicated by the other forms.<sup>29</sup>

### 12.2.5 Bodleian Library MS Douce 55

In terms of the average level of variation, the orthography reflected by the linguistic profile of MS Douce 55 (abbreviated as MS D) lies somewhere between MSS Ad and As, the average number of forms found per *LALME* questionnaire item being slightly over two and a half. Also the internal relationship of the dominant and minor forms is similar to these two texts.<sup>30</sup> Based on a general examination of the geographic distributions of the questionnaire items listed in *Table 12.5*, the text would seem to represent a Southern dialect. Forms like *nat* for ‘not’ (item 36), *lityl* for ‘little’ (item 191), and *to/too* for ‘two’ (item 275) make a northern origin very unlikely. However, there are also some features associated with northern varieties, such as the forms *thruigh/thrughe* and *thurghe* for ‘through’, which in *LALME* occur mostly in the north around Yorkshire, but do also occur sporadically in the south.

Applying the fit technique to the commonly occurring features listed in *Table 12.5* results in the cumulative exclusion template shown in *Figure 12.11*<sup>31</sup>, with the areas of occurrence for all the listed forms overlapping in the area south of London on the border of Middlesex and Surrey. *Figure 12.12* provides two alternative visualisations of the same data, which indicate that while several LPs all around the country share many of the forms attested here (many of which are very common forms occurring all over the south and Midlands), the area indicated by *Figure 12.11* does match more forms than any other area. While no single LP contains all of the forms, the strongest individual match in the area is provided by LPs 6440/5 and 6750, all localised to the same place in the southwestern corner of Middlesex. In the other areas of weaker similarity in Herefordshire and northern Essex, the closest fit is provided by LPs 7290 and 5601/2, respectively. The fact that these two LPs also exhibit significant similarity to many of the other versions of *Potage Dyvers* is a good indication of the shared dialectal features between all of the six versions, discussed in more detail in section 12.3.

If we analyse the contributions of the individual forms to the composite picture, we can see that the area south of London would seem to be minimally defined by four groups of forms: 1) the form *nat* for ‘not’ (item 36), which limits the dialect to the South and southern Midlands; 2) the forms used for ‘through’ (item 55), especially *thurghe* and *thruigh/thrughe*, which exclude much of the South and Midlands but coincide in the area south of London and in eastern Norfolk (and partially also in the central Midlands); 3) the forms *litill*, *litell* and *lityl* for ‘little’

<sup>29</sup> Considering the relatively southeastern localisation provided by the forms analysed here, it is somewhat surprising that the characteristically southeastern form *eyryn* for ‘eggs’ is used only once in this version, being overwhelmed by the 127 instances of different variants of *eggs*, which is the predominant form in the other *PD* versions.

<sup>30</sup> The average ratio between the frequency of the dominant form and the combined frequencies of the rest of the forms is roughly 3.6, lying between the 3.1 of MS As and the 4.5 of MS Ad, varying between 0.3 and 46 for individual items. All of the attested forms along with their absolute quantities are listed in the file “MSDouce55\_LALME\_profile\_quantified.xml” included in appendix G.

<sup>31</sup> Of the forms occurring more than ten times in the text, the form *yiff/yiffe* for ‘if’ and the form *togedrys/to gedrys* for ‘together’ were excluded from the analysis because they only occur in 7 and 4 LPs in *LALME*, respectively, and were not considered to provide a reliable basis for localisation.

No. ITEM form (#)
7 THEY they (17)
8 THEM hem (394)
24-20 WILL (2sg) wilt (14)
30 THEN then (146)
41 WHILE while (16)
46 NOT nat/natt (29)
55 THROUGH through/thrughe (12)
55 THROUGH thurgh/thurghe (36)
100 BUT butt (36)
37 FIRE fyre (33)
140 FLESH flesshe (15)
155 GOOD gode (39)
155 GOOD good (13)
160-20 HAVE (inf) haue (13)
191 LITTLE litel (10)
191 LITTLE litell (63)
191 LITTLE litill (10)
191 LITTLE lityl (10)
275 TWO to (12)
275 TWO too (13)
45 (S) WH- for W- x (15)
115-20 (S) DO (2sg) dost (12)
172 (S) HIS hys (33)
277 (S) UNTIL til/till (52)



Table 12.5: The forms used for analysing the dialectal ‘fit’ of MS Douce 55.

Figure 12.11: The area of occurrence shared by all of the features listed in Table 12.5, marking the most likely area of origin for MS Douce 55.

(item 191), which rule out Norfolk and Derbyshire; and 4) the form *wilt* for the second person singular of ‘will’ (item 24-20), which rules out the north of East Midlands. While the area shown in *Figure 12.11* offers the best fit in terms of the forms analysed, many of the forms used in this MS version also overlap in the central Midlands, and if we assume the forms for ‘through’ to be intruding northern forms peculiar to this text, it could also originate in the Central Midlands, around the border of Warwickshire and Leicestershire (which has a cluster of relatively closely related LPs), or even in southern Herefordshire, where the closest individual parallel is LP 7290, with also other surrounding LPs being relatively similar. However, the overall similarity of these areas is less than that of the area south of London.

Looking at the rarer forms that were excluded from the above analysis, we can see that many of them also support the attribution of the text to the area southwest of London indicated by *Figure 12.11*. For example the less frequent forms *wolt* (item 24-20), *yiff/yiffe* (item 33), *ayene* (item 37), and *renne* (item 233-10), all of which have relatively few attestations in *LALME*, all occur in this area. While *wolt* and *ayene* also occur in the Central Midlands and Herefordshire, the forms *yiff/yiffe* and *renne* do not, reinforcing the status of north Surrey or London as the most likely origin of this manuscript version.



**Figure 12.12:** A graduated dot map showing the degree of similarity between MS Douce 55 and the localised LALME profiles (darker dots representing more similar LPs), and a graduated sum of exclusion templates (darker colour indicating decreased similarity), marked with the most similar individual LPs located in the areas of greatest overall similarity.

### 12.2.6 Durham University Library MS Cosin V.iii.11 A

In terms of the average level of orthographic variation, the linguistic profile of MS Cosin V.iii.11 A (abbreviated as MS C) is quite similar to that of MS As, the average number of forms per *LALME* questionnaire item being two and a half. Similarly to MS Ad, the internal relationships of the different forms vary, although on average, the dominant forms are stronger, accounting for a greater proportion of forms than in either MS Ad or MS As.<sup>32</sup> Based on a general examination of the geographic distributions of the questionnaire items listed in *Table 12.6*, the text—like MS As—would seem to represent an Essex or East Anglian dialect.

In terms of specific dialectal forms that rule out other parts of the country as a likely origin, the forms *pei* and *hem* for ‘they’ (item 7) and ‘them’ (item 8), respectively, along with the forms *whanne* for ‘when’ (item 56) and *to* for ‘two’ (item 275) make a northern origin improbable. The forms used in this text for ‘together’ (item 268), namely *to gedere* and *to gydere* are especially typical of the Midlands and East Anglia—the latter occurring almost exclusively in East Anglia—making a Southern origin unlikely, despite the use of the forms *lite* and *lyte* for ‘little’, which are more common in the South and West Midlands but do occur sporadically also in East Anglia. Like MS As, this text also exhibits the forms *wolt* and *wilt* for the second person singular of ‘will’ (item 24-20), as well as the form *thanne* for ‘then’, also pointing to an East Anglian or East Midlands origin.

While the areas of most frequent attestation for the above mentioned items are located in various places across the South and the Midlands, the application of the fit technique reveals that the area in which the recorded attestations of all the above mentioned forms coincide most clearly is located in northern Essex, near the border of Suffolk and Cambridgeshire (*Figure 12.13*).<sup>33</sup> This would seem to agree with the preliminary localisation in Marttila (Forthcoming), which was based on less systematic examination of the linguistic profile of this MS version and located the most likely area of origin in East Anglia. As the dot map in *Figure 12.14* shows, the single LP in this area that provides the closest match to this text is LP 5601-2, which is also a close match to MSS As, D and H4016, although the LP 5591-2 in the south of the county is almost equally close. This shared similarity might indicate either a shared origin in Essex or East Anglia for these manuscripts, or—perhaps even more likely—a shared transmissional ancestry in that area. However, as the graduated exclusion map in *Figure 12.14* indicates, this text shares a number of features also with the areas of the West Midlands and the Somerset–Wiltshire area, as well as with the areas east and south of London. In these areas, the individual LPs with closest matches are LP 5651-4 for the area south of London and LP 5411-2 for Wiltshire. As with the other versions, this mixture of various general Midlands features could also indicate an origin in the capital, as was already pointed out

<sup>32</sup> The dominant forms accounting, on average, for five sixths of the forms, the ratio between the dominant and other forms extending from one third to 67. All of the attested forms along with their absolute quantities are listed in the file “MSCosinViii11\_LALME\_profile\_quantified.xml” included in appendix G.

<sup>33</sup> There is also another point in the central Midlands, at the northernmost tip of Oxfordshire, where all of the forms are found in relatively close proximity, but since it is extremely peripheral with respect to not only the southern form *lite* for ‘little’ (item 191), but also to the forms *to gydere* and *to gedere* for ‘together’ (item 268), the likelihood of this area has been judged to be marginal compared to the stronger match in East Anglia.

No. ITEM form (#)
7 THEY þei (13)
8 THEM hem (434)
30 THEN þan (81)
30 THEN þenne (62)
30 THEN þen (58)
30 THEN than (13)
30 THEN thanne (12)
41 WHILE while (21)
52 THERE þer (11)
55 THROUGH thorow (100)
56 WHEN whan (47)
56 WHEN whanne (12)
100 BUT but (40)
137 FIRE fire (61)
155 GOOD good (42)
155 GOOD goode (29)
191 LITTLE lite (43)
191 LITTLE litel (17)
191 LITTLE lyte (16)
268 TOGETHER to gedere (52)
268 TOGETHER to gydere (21)
275 TWO to (21)
67 (S) PPL PREFIX i/I (12)
115-20 (S) DO (2sg) dost (11)
277 (S) UNTIL til (46)
277 (S) UNTIL tyl (11)

**Table 12.6:** *The forms used for analysing the dialectal ‘fit’ of MS Cosin V.iii.11 A.*



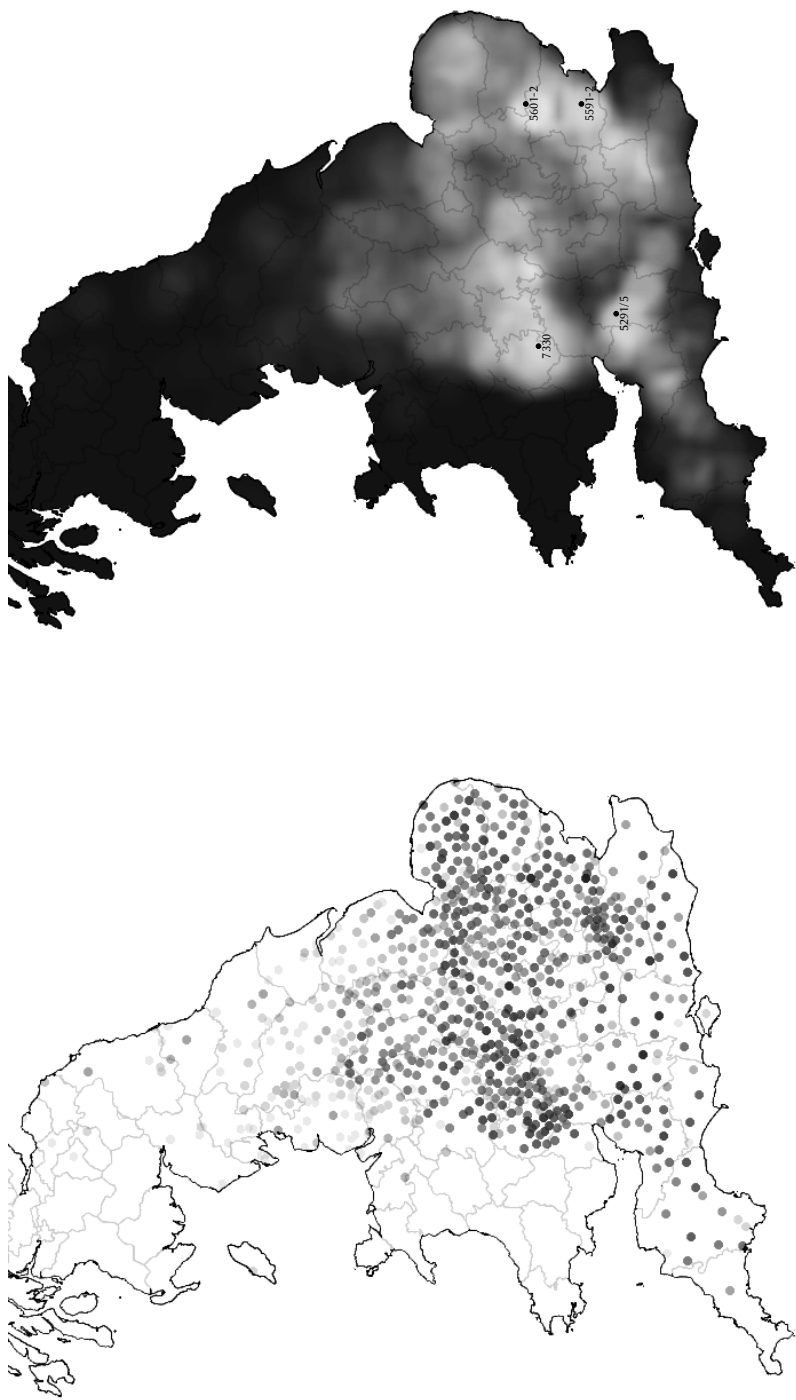
**Figure 12.13:** *The area of occurrence shared by all of the features listed in Table 12.6, marking the most likely area of origin for MS Cosin V.iii.11 A.*

above.

If we analyse the contributions of all the listed forms to *Figure 12.13* (the exclusion maps for each of the items can be found in appendix G), we can see that the set of features that would minimally seem to limit the area of origin to the border area between Essex, Suffolk, and Cambridgeshire consists of a combination of the forms of ‘then’, especially *thanne* and *thenne*, the forms of ‘little’ (especially *lite*<sup>34</sup>), and the forms used for ‘together’, i.e. *to gydere* and *to gedere*. All of the other significant features analysed completely overlap with the area delimited by these forms, and their inclusion in the analysis serves to merely strengthen the localisation.

Although the major forms would seem to coincide quite neatly in the Essex–East Anglia area, even a cursory look at the forms that are rarely attested either in the text or in *LALME* reveals a variety of influences also from outside this area, especially from the central Midlands. For example the item ‘these’ (item 2) occurs in the text three times, always in a different form. The form *this* is neutral and is

<sup>34</sup> The form *lityl* is also strongly associated with East Anglia, but it has not been included in the analysis as it only occurs nine times in the text.



**Figure 12.14:** A graduated dot map showing the degree of similarity between MS Cosin V.iii.11 A and the localised LALME profiles (darker dots representing more similar LPs), and a graduated sum of exclusion templates (darker colour indicating decreased similarity), marked with the most similar individual LPs located in the areas of greatest overall similarity.

common in several areas around the country, but the forms *peys* and *peyse* appear only in four LPs in a narrow vertical strip along the central Midlands, extending through Leicestershire, Northamptonshire and Buckinghamshire. Similarly, the form *ilche* for ‘each’ (item 12) is recorded only in four LPs in *LALME*, three of them situated in Warwickshire and one in northern Norfolk, indicating central Midland influence. These influences may indicate that either the manuscript itself or its scribe originated there, the eastern features being inherited from an earlier exemplar or from a variety learned later in the scribe’s professional career. However, many of the minor forms excluded from the analysis either because of their low frequency in the text or their scarce attestation in *LALME*, also occur in Essex. For example the occurrence of *-eng-* in place of ‘-ing-’, mainly in the word ‘wing(s)’ (*weng(es)*), is attested predominantly in East Anglia and northern Essex, while the forms *wilt* and *wolt* used for the second person singular of ‘will’ occur equally in Essex and in the West Midlands. Also forms like *aftur* (item 29) and *whyle* (item 41) which occur predominantly in other parts of the country (in the West Midlands and the South) have sufficiently frequent attestations in Essex to make them plausibly native to that area.

## 12.3 Shared features

Based on the linguistic profiles of the six *PD* versions, it is apparent that in addition to dialectal differences, they—not altogether unexpectedly—also exhibit considerable linguistic similarity. The two most obvious sources for linguistic similarity between the six versions are their shared textual ancestry and the general process of linguistic standardisation taking place in English over the 15<sup>th</sup> century. The dialectal profiles of the six versions will here be examined from both of these viewpoints, focusing both on the degree of similarity between individual pairs of versions, and on the kinds of features that are shared by all of them and their geographical distribution. With regard to the first question, I will compare each pair of collections individually to determine the extent to which their dialectal profiles coincide, and quantify this similarity based not only on the number of shared forms but also on their primacy in the profile, i.e. whether they are major, minor or negligible forms for the item in question. With regard to the second question, I will examine the geographical distributions of those forms that are shared by all six versions—whether as major, minor or negligible forms—to determine whether they point towards a specific geographic region, indicating the influence of a shared ancestor, or whether they are neutral and geographically undifferentiated forms, pointing towards the influence of standardisation.

### 12.3.1 Dialectal similarity between versions

Since the selection of *LALME* questionnaire items exhibited by each manuscript version is somewhat different and the focus here is on differences in the *forms* employed, the quantification of similarity between the manuscript versions must be based on an aggregation of the similarities between the forms exhibited for individual *LALME* questionnaire items by the two texts. Since we are not interested

in lexical similarity but rather in orthographical similarity,<sup>35</sup> only questionnaire items present in both of the versions being compared are taken into account. As the aggregate similarity score is based on the average of the similarity scores of the shared questionnaire items, differences in the number of items shared by different pairs of versions are not a problem. The similarity score will also need to be normalised in terms of the varying number of forms for different questionnaire items, but will still need to reflect any differences in the number of forms exhibited for a given item by the two versions to be compared. Furthermore, the score should reflect not only differences in the forms used by the two versions, but also differences in the relative frequencies with which these forms are used, which means that differences in more frequent forms should have a larger negative impact on similarity than differences in infrequent forms. All of this means that the similarity score for two versions exhibiting the same forms for an item with the same relative frequencies (or ‘primacy levels’<sup>36</sup>) should be 1, and the score for pairs with no shared forms should be 0. Since we are comparing decontextualised aggregate values which are not dependent on a version-specific context, the score should also be symmetrical, i.e. the same for each of the two texts compared. In order to fulfill these requirements, the following formula is used to calculate a similarity score for each *LALME* questionnaire item shared between a pair of *PD* versions, with Table 12.7 providing examples of the values produced by this formula for different cases:

$$\frac{3(f_3^0 + \frac{2}{3}f_3^{-1} + \frac{1}{3}f_3^{-2}) + 2(f_2^0 + \frac{2}{3}f_2^{-1} + \frac{1}{3}f_2^{-2}) + (f_1^0 + \frac{2}{3}f_1^{-1} + \frac{1}{3}f_1^{-2})}{(3f_3 + 2f_2 + f_1)}$$

where

$f$	number of distinct dialectal forms (counted separately for each version)
$f_3$	number of major forms (primacy level 3, no parentheses)
$f_2$	number of minor forms (primacy level 2, single parentheses)
$f_1$	number of negligible forms (primacy level 1, double parentheses)
$f^0$	number of forms matched on the same primacy level in the other version
$f^{-1}$	number of forms matched one primacy level apart in the other version
$f^{-2}$	number of forms matched two primacy levels apart in the other version

E.g.

$f_2^{-1}$	number of minor forms matched one primacy level apart in the other version
------------	--

While more complicated than a straightforward percentage of forms shared by two versions of the text, this formula has the benefit of also accounting for differences in relative frequency, which is here argued to be a significant part of the linguistic profile of a text. While a straightforward unquantified comparison of forms would find no difference in the forms used for ‘it’ (item 6) in MSS As and

<sup>35</sup> Although in many cases, the same item may also be represented by different lexical tokens in different dialects.

<sup>36</sup> The concept of *primacy levels* is here used to refer to the classification of the relative frequency of a form in *LALME* using either single or double parentheses. Forms with no parentheses or *major* forms are considered to have a primacy level of 3, forms with single parentheses (minor forms) a primacy level of 2, and forms with double parentheses (negligible items) a primacy level of 1.



	Example 1	Example 2	Example 3	Example 4	Example 5	Example 6
LP 1 forms	A, B, (C)	A, B, C	A, (B), ((C))	A, (B), ((C))	A, ((B))	A, B, (C)
LP 2 forms	A, B, (C)	B, C, D, E	A, B, ((C))	C, (B), ((D)), ((E))	B, ((A))	D, (E), ((F))
Formula	$\frac{12+4+0}{16}$	$\frac{12+0+0}{21}$	$\frac{8+1\frac{1}{3}+2}{13}$	$\frac{1+4+\frac{1}{3}}{13}$	$\frac{2+0+\frac{2}{3}}{8}$	$\frac{0+0+0}{14}$
Value	1.0	0.57	0.87	0.41	0.33	0.0

**Table 12.7:** Examples of similarity scores calculated for the forms of a single questionnaire item in two linguistic profiles, the capital letters representing dialectal forms and the parentheses their relative frequencies as per *LALME* conventions.

D, namely *it* and *hit*, the quantitative reality is quite different, the former form accounting for 89 per cent of the total in MS D but only 16 per cent in MS As, while the latter in turn cover 84 per cent of the forms in MS As but only 9 per cent in MS D, meaning that in the vast majority of cases, they in fact employ different forms.<sup>37</sup> While the issue of the degree to which the quantitative differences should be considered to affect the dialectal similarity of texts is a difficult one, the approach taken here is based on the simple relative quantification used in *LALME*.

The multipliers for the different primacy levels are based on the factor of three used to define the *LALME* quantification (*LALME*: vol.3: xiv, see also section G.1); since the frequencies of ‘minor’ and ‘negligible’ forms (marked with single and double parentheses in *LALME*) are below one third and below two thirds, respectively, of the frequency of the dominant form or less, their influence on the similarity metric has been scaled down by the same factors. The use of the same factor of three for quantifying the effect of *differences* in frequency is a somewhat more controversial solution, but has here been adopted both for its simplicity and the rather straightforward logic that a form occurring half as often (in relative terms) can be considered to represent half the similarity of a form occurring equally often. The use of the *LALME* method of quantification also has the benefit of making it compatible with the data contained in *LALME* itself—while a metric based on the absolute frequencies of the forms might make more accurate distinctions and be more straightforward to implement, the method adopted here allows direct comparisons to be made to any LP recorded in the *LALME*.

After having been calculated for each shared questionnaire item of each pair of versions, the individual similarity scores were averaged over each pair, the average scores being presented in *Table 12.8*, together with the standard deviation and the number of shared items for each pair is. The fact that neither the similarity score nor the degree of its variation (measured by the standard deviation) seem to correlate with the number of items shared by a version pair would seem to predict that the formula scales well, allowing pairs of texts sharing a widely varying number of items to be compared.

As can be seen in *Table 12.8*, the differences in the aggregate dialectal similarity scores between the different pairs of versions are quite small, ranging from the minimal similarity of 33.6 per cent between MSS H279 and Ad to the maximal

<sup>37</sup> According to the formula used here, the similarity for this case—which corresponds to Example 5 in *Table 12.7* is 0.33 instead of the 1 given by unquantified comparison of forms.

	Harley 279	Additional 5467	Ashmole 1439	Douce 55	Cosin V.iii.11
Harley 4016	A = 38.7% $\sigma$ = 38.1pp # = 123	A = 42.5% $\sigma$ = 37.8pp # = 118	A = 44.5% $\sigma$ = 38.7pp # = 124	A = 42.5% $\sigma$ = 37.8pp # = 116	A = 47.2% $\sigma$ = 39.2pp # = 115
Harley 279		A = 33.6% $\sigma$ = 35.4pp # = 117	A = 44.3% $\sigma$ = 36.5pp # = 140	A = 38.5% $\sigma$ = 36.3pp # = 111	A = 47.8% $\sigma$ = 36.5pp # = 108
Additional 5467			A = 42.4% $\sigma$ = 33.3pp # = 117	A = 39.3% $\sigma$ = 35.3pp # = 116	A = 40.2% $\sigma$ = 35.8pp # = 113
Ashmole 1439				A = 39.6% $\sigma$ = 35.6pp # = 114	A = 49.5% $\sigma$ = 38.7pp # = 112
Douce 55					A = 36.3% $\sigma$ = 35.7pp # = 110

**Table 12.8:** *The arithmetic mean (A) and standard deviation ( $\sigma$ ) of the LALME questionnaire item similarity scores, along with the number of shared items (#), for each manuscript pair.*

similarity of 49.5 per cent between MSS As and C, the average level of dialectal similarity between the different versions, calculated according to the formula described above, being 41.8 per cent. As the extremely high figures for the standard deviation for the similarity scores indicate, the average level of similarity reflects mostly the proportion between items with a very high and very low similarity, rather than of the typical match level of an individual item. In terms of their average dialectal distance from all of the other versions, the most dialectally idiosyncratic version is MS D, whose average level of similarity to the other versions is 39.2 per cent, while the version that is least different from the others and can thus be considered to be the most characteristic of the shared dialectal features is MS C, with an average overall similarity score of 44.2 per cent. Comparing the dialectal similarity scores to the structural similarity of the versions examined in chapter 13 below, we can see that the dialectal similarity of the different versions does not seem to correlate with their structural similarity. This would seem to indicate that even the faithful preservation of the textual structure of the collection by the scribe does not imply the preservation of its dialectal features. Since none of the manuscript versions are sufficiently similar dialectally to be considered linguistically faithful copies of either each other or of a common ancestor, the converse—i.e. whether a linguistically faithful copy also tends to preserve textual structure—cannot be evaluated based on this data. Together, these observations would seem to indicate that in medieval practical texts like recipes, textual structure is to be considered a more integral component of textual identity—even in discourse colonies where it is more fluid than in mainstream texts—than its dialectal or linguistic characteristics.

While the exploratory method employed here does provide a rough idea of the overall similarity of entire collections, it does not provide any indication of the degree to which this overall similarity represents the typical similarity of an individual pair of parallel recipes. For further research, this method could be refined to focus on the similarity of individual pairs of parallel recipes, examining the

degree to which the dialectal similarity varies within the collection and whether it can reveal specific sets of recipes that have an especially close or distant dialectal affinity to their parallels in other versions of the collection. Another development of this method to be undertaken in a future research article would be to go beyond the absolute comparison of forms and extend the graduated approach to within the forms themselves, taking into account the degree to which two dialectal forms differ from each other by using a suitable edit distance metric such as Damerau-Levenshtein (Damerau 1964; Levenshtein 1966) or Jaro-Winkler (Jaro 1989; Winkler 1990) distance. This would allow us to properly quantify the difference between relatively weakly contrasting pairs like *well/welle* and more significant differences like *enow/ynoghe*.

### 12.3.2 Characteristics of shared forms

Despite their considerable differences leading to the relatively low similarity scores seen above, there are also a number of dialectal forms occurring in all six versions of the *PD*, representing a kind of ‘lowest common denominator’ for the dialects of the different versions. The majority of these forms, which are listed in *Table 12.9*, would seem to be dialectally unmarked ‘neutral’ forms that are in general use over most of the area covered by *LALME* and on their way to becoming a part of the standard variety of English. Unsurprisingly, the items attested in every version consist mostly of grammatical words and other extremely common lexical items, supplemented by register-specific items like ‘flesh’, ‘fire’, and ‘good’.<sup>38</sup> There are, however, some shared forms that have a marked—and in some cases quite an exclusive—dialectal distribution. Applying the ‘fit’ technique—using the procedures described above—to these shared forms results in the cumulative exclusion map shown in *Figure 12.15*.<sup>39</sup> The most significant contributors to the map, i.e. the forms with the most limited geographical distribution in *LALME*, are the form *wolt* for the second person singular of ‘will’ (item 24-20), which occurs sporadically throughout the South, East Anglia and the southern West Midlands, with the largest concentration of occurrences along the Welsh border; initial *w-* for *wh-* (item 44), which also occurs in texts scattered around the Midlands, East Anglia and the Southwest; and the form *flesshe* for ‘flesh’ (item 140), which occurs mainly in the East Midlands and the North, extending to the Southwest along the Welsh border, with a separate enclave in the London–West Essex area.

While it would be tempting to try and associate this map with the area of origin for the ‘original’ version or *archetype* of the collection, or even a shared intermediate exemplar—should one even exist—it is more likely to represent simply the intersection or ‘lowest common denominator’ of the different Southern and Mid-

<sup>38</sup> While the adjective “good” does not seem specific to the culinary recipe register, it is one of the most common adjectives used in the *Potage Dyvers* collection, second only to ‘fair’ and can be considered very characteristic of medieval recipes.

<sup>39</sup> Some of the forms shared by all the *PD* versions were not included in the analysis since the items represented by them were only included in the northern version of the questionnaire and not judged to be diagnostically relevant for the southern part of England. These forms have been marked with the identifier (*N*) in *Table 12.9* and grouped together at the end of the list. Other forms, marked with an asterisk (\*) have been excluded because the items they represent are not sufficiently attested in *LALME* and their geographical distribution could thus not be reliably determined. All of these forms excluded from the analysis have been printed in grey in *Table 12.9*.

ITEM	Harley 4016	Harley 279	Additional 5467	Ashmole 1439	Douce 55	Cosin V.iii.11 A
6 IT	((it))	it	((it))	((it))	it	it
7 THEY	they	((they))	they	((they))	they	((they))
8 THEM	hem ((hem))	hem ((hem))	hem ((hem))	((hem)) ((hem))	hem ((hem))	hem (hem)
14 MAN	man	man	man	man	man ((man))	man
19 IS	is	is	is	is	is	is
21 WAS	was	was	was	was	was	was
24-20 WILL 2sg	((wolt))	wolt	(wolt)	wolt	((wolt))	(wolt)
28 FROM	fro	fro	fro	fro	fro	fro
29 AFTER	after	after	after	after (after)	after	(after)
30 THEN	þen	(þen)	((þen))	((þen))	((þen))	þen ((þen))
44 WH-	wh-	wh-	wh-	wh-	wh-	wh-
44 WH-	((h-))	((h-))	((h-))	((h-))	((h-))	((h-))
46 NOT	not	(not)	(not)	not	((not))	not
56 WHEN	((when))	((when))	when ((when))	((when))	when (when)	((when))
57 Sb pl	-es ((-es))	((-es)) ((-es))	-es ((-es))	-es ((-es))	-es	-es (-es)
57 Sb pl	((-ys))	-ys	-ys	-ys	((-ys))	-ys
57 Sb pl	((-is))	((-is))	(-is)	((-is))	((-is))	((-is))
100 BUT	but	but	((but))	but	((but))	but
137 FIRE	((fyre)) ((fyre))	fyre ((fyre))	((fyre))	(fyre)	((fyre))	((fyre)) ((fyre))
140 FLESH	flesshe	(flesshe)	flesshe flesshe	flesshe	flesshe	((flesshe))
155 GOOD	good	((good))	(good)	((good))	(good)	good
160 HAVE pres	haue	haue	haue	haue	haue	haue
*172 (S) HIS	his	his	his	his	(his)	his
221 OR	or	or	or	or	or	or
275 TWO	((to))	to	((to))	to	to	to
277 S UNTIL	til	((til))	til	til	((til))	til
*312-10 -ER cpv	-er	-er (-er)	-er	-er -er	-er ((-er))	-er (-er)
1 (N) THE	the	((the))	the	((the))	the	((the))
1 (N) THE	þe	þe	((þe))	þe	((þe))	þe
18 (N) WERE	were (were)	were (were)	were	were	were	were
26 (N) TO prep	to	to	to	to	to	to
27 (N) TO inf	to	to	to	to	to	to
38 (N) ERE	or	or	or	or	or	or
65 (N) Weak ppl	-ed	((-ed))	(-ed)	-ed	-ed	((-ed))
66 (N) Str ppl	-e ((-e))	-e	-e -e	-e	-e ((-e))	-e
66 (N) Str ppl	(-en)	((-en))	((-en))	((-en))	(-en) ((-en))	((-en))
66 (N) Str ppl	((-yn))	((-yn))	((-yn))	((-yn))	((-yn))	((-yn))
71 (N) ABOVE	aboue	aboue	aboue	((aboue))	(aboue)	aboue
84 (N) BE inf	be	be	be	be	be	be
102 (N) BY	by	(by)	by	by	by	((by))
112 (N) DAY	day	day	day	day	day	day
199-10 (N) MAY 1/3 sg	may	may	may	may	may	may
213 (N) NEW	newe	newe	newe	(newe)	newe	newe
220-20 (N) ONE pron	on	on	on	on	on	on
222 (N) OTHER	other	((other))	other	((other))	(other)	other

Table 12.9: LALME forms shared by all six MS versions of Potage Dyvers.



**Figure 12.15:** The area of occurrence for the features shared by all six MS versions of the Potage Dyvers collection listed in Table 12.9.

**Figure 12.16:** A cumulative exclusion map combining the graduated exclusion maps for all the six versions of the Potage Dyvers collection, with the lighter areas representing the overlapping portions of the most likely areas of occurrence for each version.

lands dialects used for the different manuscript versions. This can also be seen in *Figure 12.16* which combines the graduated exclusion maps of the six versions to reveal that the areas delimited by the shared features are also the ones least excluded by the forms particular to individual collections. Thus the most that can be concluded from these shared forms and their distribution is that at least in the mid-to-late 15<sup>th</sup> century the *Potage Dyvers* collection seems to have been copied and circulated mainly in dialects associated with the southern part of the country, most likely in Essex and the area around London, and in the area at the intersection of the West Midlands and the Southwest near the Welsh border.

Whether these similarities in the dialectal features of the six versions of *Potage Dyvers* are due to the influence of relict forms from a shared ancestor or to the more general process of dialectal standardisation is of course impossible to say with certainty. However, based on the nature of the forms shared by all the *PD* versions and the area in which they co-occur, combined with the relatively late date of these manuscripts and the observation that practical texts copied for a local audience usually tend to get quite freely translated into a local dialect, it seems more likely that we are in fact dealing with the gradual disappearance of dialectal differences rather than any coherent underlying influence.

## 12.4 Variety and degree of standardisation

The significant internal variation within the individual versions—demonstrated by the occurrence of a wide variety of forms for many of the *LALME* items—should not be surprising considering the nature of the collection as a *discourse colony* and the potential disparity in the textual histories of individual recipes that it entails. In addition to containing several levels of dialectal use because of multiple rounds of copying as described by Benskin and Laing (1981), the differences in the order and selection of recipes contained in the different versions also hint at the possibility that they may have been compiled from multiple sources, different parts of the collection representing different dialectal backgrounds.<sup>40</sup> This means that these texts can be seen to represent a kind of linguistic composite described by Benskin and Laing (1981) in two different dimensions: the ‘depth’ of the textual tradition and the ‘length’ of the recipe sequence. Thus the caveat made by McIntosh, Samuels and Benskin (*LALME*) in connection with the linguistic profiles described in *LALME* applies to an even greater degree to the linguistic profiles of the *Potage Dyvers* manuscripts:

A scribal profile is here less a characterisation of an individual writer’s spontaneous usage than a statement of his linguistic tolerance. The range and relative frequency of variant forms, since they are text-determined, cannot be held to represent the usage of any one writer, but only of an indeterminate sample from the literate community to which he belonged. (LALME: vol.3, x)

In a similar vein, Milroy (1992) has argued that the occurrence of forms ascribed to several different dialects can indicate either that “the text is composite and has been copied by scribes from different dialect areas” (132) (as discussed e.g. by Benskin and Laing 1981, *LALME* and Laing 1989), or that all of the forms were current in the dialect of the scribe or the author, “or—more properly—of the speech community to which he belonged” (Milroy 1992: 132). In the case of the *PD* manuscript versions, all of which seem to have been written by a single scribe (apart from some additions and corrections which have here been excluded from the dialectal analysis), and were clearly not copied verbatim—as is witnessed by the differences in their dialectal features and even syntactic and textual structure (see the parallel reading edition in appendices C and D)—there is in fact a strong case for arguing that whatever forms are found in the manuscript can be considered to belong at least to the passive repertoire of a single scribe.<sup>41</sup> This does not, however, mean that the *PD* texts would not contain influences from several different dialects, merely that these influences are likely to be relatively subtle, as in the case of the forms of ‘not’ in MS H279 discussed above.

<sup>40</sup> The different hypothetical sources and the groups of recipes likely to be derived from them are examined in chapter 13.

<sup>41</sup> Since the scribe would have a very low threshold for replacing ‘foreign’ forms with his own habitual ones, the occurrence of seemingly dialectally disparate forms could be taken as an indication that they were considered acceptable—if not spontaneous—by the scribe.

### 12.4.1 Distribution of variant forms

While dialectal variation resulting from multiple rounds of copying is difficult to detect, variation due to different parts of the text having been copied from different sources is—at least in theory—more easily detectable. While not distinguishable in the aggregate analysis of the entire collection, it should be detectable as variation between different parts of the collection. While the scope of this thesis does not allow for the creation and comparison of individual LPs for each of the recipes in each of the versions,<sup>42</sup> the effects of the fragmented nature of a discourse colony can nevertheless be approached by looking at the distribution of parallel dialectal forms across the length of the recipe collection. Whereas their even distribution throughout the collection implies either consistent influence from a single exemplar—or multiple ones written in the same dialect—or the natural alternation of forms native to the speaker's idiolect, their exclusive or increased occurrence in specific parts of the collection implies the possibility that the part in question has been transmitted through exemplars written in a different dialect.

The dialectal features selected for the exploratory analysis presented here were chosen individually for each version based on their linguistic profile. In order to find the forms that can most reliably be used to detect local variation within the recipe collection, the quantitative linguistic profiles of each version were mined for items that a) have a dominant form occurring at least ten times in the text of the version, and b) have at least one secondary form whose frequency is at least one third of the dominant form. These criteria resulted in the selection of 9–14 items for each of the versions, most of which fulfil the criteria in more than one version but none in all of the versions. It is important to note that the items were selected on purely quantitative grounds without reference to their significance as dialectal markers for the dialectal origin established in section 12.2.<sup>43</sup> In terms of the forms attested for these items, only those forms with a frequency of at least one third of the dominant form were included in the analysis.<sup>44</sup> *Figure 12.17–Figure 12.21* present the items and forms selected for each text according to the above mentioned criteria and indicate the sequential distribution of the different forms in terms of the individual recipes included in the collection.<sup>45</sup>

While differences in the distributions of the parallel forms of individual items do not as such indicate divergent textual histories for different parts of the collection, the coincidence—or near-coincidence, as will be argued below—of changes in the forms use for several different items, especially if it also coincides with textual divisions of related recipes, can be seen as evidence of diverse sources. The

<sup>42</sup> This method of analysis would also be problematic due to the limited length of the recipes and requires further investigation as to its feasibility.

<sup>43</sup> This means that even items included only in the northern version of the *LALME* questionnaire were included here since even if not diagnostically relevant for the dialect of the current version, they can nevertheless reveal traces of underlying influences from earlier exemplars potentially written in different (even northern) dialects.

<sup>44</sup> While the occurrence patterns of dialectally distinctive minor and negligible forms would also be potentially interesting—and possibly a stronger indicator of localised dialectal influence within the collection—both the manual selection of these forms and the visualisation and analysis of a much larger number of both items and forms would involve a considerable amount of labour and is beyond the scope of the present exploratory analysis.

<sup>45</sup> It should be noted that the figures indicate merely the presence or absence of the form in a recipe and do not quantify the number of occurrences within a single recipes.

following paragraphs describe the discernible patterns (or their absence) in the figures for each of the *PD* versions, relating them to the internal structure of the collections.

### MS Harley 4016

For MS H4016, the items fulfilling the above mentioned criteria would seem to be mostly neutral forms not exclusively associated with any local variety.<sup>46</sup> In terms of their distribution within the text, the forms used for most of the items—e.g. ‘the’, ‘then’, ‘all’, ‘cast’, ‘good’, ‘little’ and ‘thou’—are clearly equal components of the scribe’s active repertoire, as they are used equally throughout the collection, with all forms frequently occurring within a single recipe. However, there are some forms, namely the forms *pei* and *thei* for ‘they’ (item 7), *-yng* for the present participle suffix (item 58) and the forms of ‘well’ (*well/wel*, item 281), that exhibit interesting patterns in their occurrence.

The first of these is a break somewhere between recipes 106 and 110, the former being the last recipe in which either of the forms *pei* or *thei* occurs in the collection and the latter the first recipe in which the form *-yng*—which is the form with wider currency in the *LALME*—occurs for the present participle suffix. In terms of the content of the recipes, this break coincides with a move from recipes for fish dishes (ending at recipe 107) to (almond) milk based pottages and custards, making it possible that the scribe (of either the current MS or its exemplar) changed exemplars at this point, although the different forms do not seem to imply a dialect difference between the exemplars but are more likely to result from the personal preferences of their scribes. The other interesting pattern is not as clear-cut and involves a more complicated vacillation between different combinations of forms for the items ‘well’ (AV),<sup>47</sup> some of which seem to correlate with the general trend in the forms of ‘enough’, *ynogh* predominating in the first half of the collection and *ynowe* in the second (with the shift occurring around recipe 80). While the linguistic shift here is less noticeable than in the previous case, recipe 80 does mark a semantic shift from a section of meat dishes to one of non-meat pottages based on bread and eggs.

### MS Harley 279

For MS H279, the items fulfilling the above-mentioned criteria are on average more dialectally marked than those for MS H4016. Although their areas of occurrence

<sup>46</sup> Of the forms included here, *the* and *all* are slightly more frequent north of Humber than *pe* (item 1) and *al* (item 75), although their areas of occurrence overlap for the most part. The form *caste* (item 106) does not occur north of Yorkshire, while the form *ynowe* (item 125) is much rarer in *LALME* than the other major form *ynogh* and occurs mostly in the Midlands, whereas the form *well* is less attested in the Midlands and East Anglia than *wel*, which is especially characteristic of the latter. The forms for the rest of the items—‘they’, ‘then’, ‘good’, ‘little’ and ‘thou’ are not dialectally distinctive.

<sup>47</sup> The form *well* is used exclusively until recipe 33 and the form *wel* from recipe 38 to recipe 79, while the section from recipe 80 to recipe 123 uses both of the forms, with both forms occasionally (recipes 100 & 102) occurring in the same recipe. From recipe 125 onwards, the form *well* is again used exclusively, apart from a single occurrence of *wel* in recipe 174. However, the relatively low frequency of this item, especially for the initial part of the collection, as well as the lack of dialectal differentiation between the forms means that the pattern might also be purely coincidental.



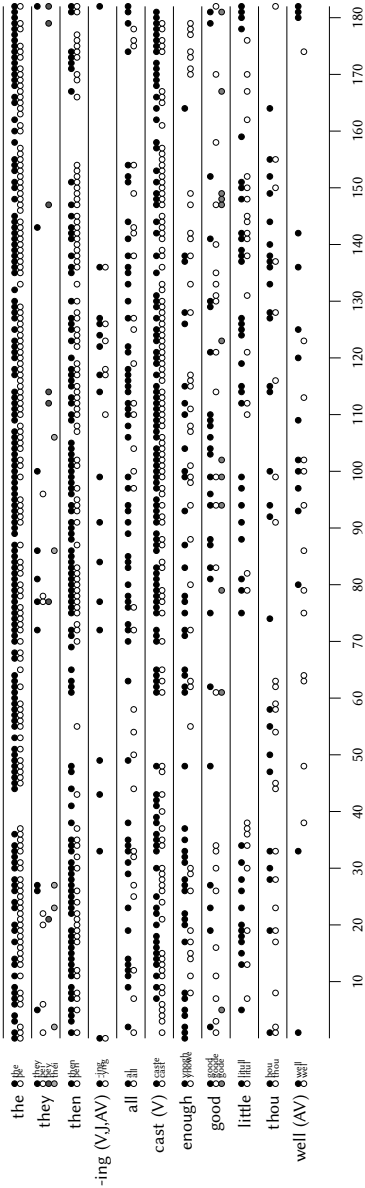


Figure 12.17: The distribution of parallel forms for selected LALME items in terms of the recipes included in MS Harley 4016.

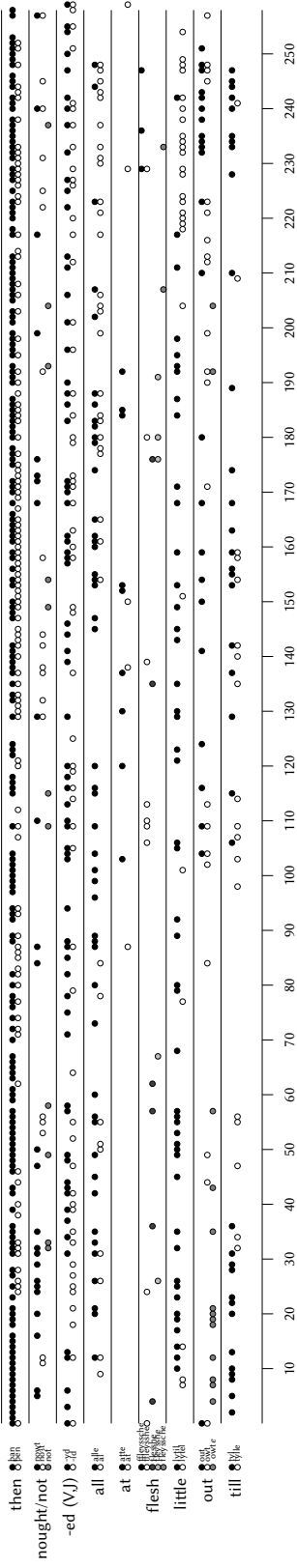


Figure 12.18: The distribution of parallel forms for selected LALME items in terms of the recipes included in MS Harley 279.

overlap for the most part in the Midlands area, many of them exhibit differing emphasis either in the north-south or east-west direction.<sup>48</sup> While some of the forms included in *Figure 12.18*—such as those for ‘then’ (item 30), ‘not’ (item 46), ‘at’ (item 82), and the weak past participle suffix (item 65)—occur throughout the collection with no distinguishable large-scale patterns, there are also several forms that exhibit some degree of spatial clustering within the collection.

The most noticeable of the patterns occurring in this text include the clear shift from *lytil* to *lytel* for the final part of the collection from recipe 218 onwards,<sup>49</sup> and the similar shift from *flesshe/fleysshe* forms of ‘flesh’ (item 140) to *fleyssche* from recipe 207 onwards. The *-sche* form does not occur before this point in the collection and the *-she* forms only occur once after it. Since these shifts correspond to the break between the subcollections for “leche vyaundez” and “bake metis”, beginning at recipe 218, they would seem to point towards the two subcollections having been copied from different exemplars at some point during their textual history. Other notable patterns include the exclusive use of the form *alle* for ‘all’ in recipes 87–147, followed by a gradual shift, from recipe 154 onwards, to *al*, which becomes clearly dominant from recipe 200 onwards.<sup>50</sup> Like the previous shift, also the shift from *alle* to *al* coincides with a textual break between the main part of the collection and the named subcollection for “leche vyaundez” which begins at recipe 154. The gradual nature of the shift between the forms could well result from the scribe gradually adapting his use from the exclusive use of *alle*, either his own preference or that of his exemplar for the main part of the collection, to that of another exemplar used as a source for the first subsection.<sup>51</sup> While the exclusive use of the form *owt/owte* in exclusion of *out* in recipes 4–101 and the converse in recipes 116–168 does not seem to correspond to any obvious structural or culinary division in the collection, these forms also seem to roughly follow the borders of the two subcollections, the form *owt/owte* being predominant in recipes 171–221, after which *out* becomes the clearly dominant form. The most complicated of the discernible patterns is the vacillation between the forms *tyl*

<sup>48</sup> For example the form *þan* for ‘then’ (item 30) seems to be more common in the east of the country and *þen* in the west. Similarly, the form *nowt* for ‘not’ (item 46, treated separately as ‘nought’ in the present edition) occurs only in East Anglia, whereas the form *tylle* for ‘until’ (item 277, treated separately as ‘till’ in the present edition) occurs all over the South and West Midlands but *not* in East Anglia. A more subtle difference is observed with the forms for ‘little’ (item 191), of which *lytil* is more common in East Anglia and the Midlands, while *lytel* occurs more frequently in the South and Midlands. In terms of the north-south direction, the form *al* is more frequent north of the Humber and on the coast of East Midlands than *alle* (although they seem to be equally frequent in the West Midlands and East Anglia), and the area of occurrence of the forms *owt* and *owte* (item 225) does not seem to extend to the Cambridgeshire–Huntingdonshire–Bedfordshire area in the south, unlike that of *out*, which covers the entire area of the northern LALME questionnaire. Of the other items, the forms of the weak past participle (item 65) and the forms for ‘at’ (item 82) do not exhibit differences in their dialectal affiliation, while the forms used for ‘flesh’ (item 140), namely *fleysshe* and *fleyssche* do not really occur in the LALME, the former occurring only in the profile for this text and the latter being entirely absent.

<sup>49</sup> The form *lytel* does also occur elsewhere in the collection, but as a clear minority form, while the form *lytil* only occurs once after recipe 218 (in recipe 242, which is also the second recipe in which both of the forms co-occur).

<sup>50</sup> Before recipe 87, the form *al* occurs relatively infrequently as a minority form—occasionally co-occurring with *alle*.

<sup>51</sup> This kind of gradual shift would imply a copying style somewhere between the active translation of the source to the scribe’s own dialect and the faithful copying of the original dialect.

and *tylle* for ‘until’ (item 277). The former form is used exclusively until recipe 31 but is replaced by the latter, which in turn occurs exclusively in recipes 47–103, after which they both occur, sometimes within a single recipe, until recipe 159, after which *tyl* clearly predominates, *tylle* occurring only twice. Of these shifts, the last corresponds approximately to the beginning of the “leche vyaundez” subsection, while the beginning of the co-occurrence of both forms would seem to correspond to the shift from *owt/owte* to *out* described above, possibly indicating a change of exemplars—or a shift of forms in the exemplar—somewhere between recipes 100 and 110. Unfortunately, none of these forms are sufficiently exclusive in their regional affiliation to establish dialectal identities for the postulated exemplars beyond that established for the whole collection in section 12.2.

### MS Additional 5467

For MS Ad, only few of the forms included in this analysis and listed in *Figure 12.19* are dialectally distinctive,<sup>52</sup> and many of the pairs or groups of forms that fulfil the formal criteria presented above can be considered to be completely interchangeable, as is indicated by their undifferentiated usage patterns.<sup>53</sup> In addition to barely distinct pairs, there are also several forms that are frequent in this text but occur in very few—often scattered—*LALME* profiles or do not occur in the *LALME* at all.<sup>54</sup> However, some forms which do not seem to be dialectally distinct nevertheless seem to differentiate between different parts of the collection. These kinds of forms include the forms for the weak past participle suffix (item 65), the form for ‘out’ (item 225), and the forms *til* and *till* for ‘until’ (item 277).

Perhaps the clearest patterns here, as in the other versions, are clear binary divisions where a specific form occurs only in the beginning or end of the collection. This kind of pattern is exhibited by the form *not* (item 46), which occurs only up to recipe 61, and the form *tyll* for ‘until’ (item 277), which only occurs up to recipe 76, and to an extent also by the forms *whanne* (item 56) and *keste/keste*, which occur only up to recipe 69 (*whanne*) and 100 (*keste*), apart from single occurrences in recipes 179 and 137, respectively. Although these breaks do not seem to coincide with any text-structural divisions or changes in the content of recipes, the first three do coincide roughly with each other and do not cut across any coherent

<sup>52</sup> Forms which do have some level of dialectal association include *nat* for ‘not’ (item 46) and *whanne* for ‘when’ (item 56), both of which are more restricted in their areas of occurrence than their parallel forms, attested in *LALME* only for the southern Midlands and East Anglia. The form *litill* (for ‘little’, item 191) excludes both East Anglia and the Southwest, and *lite* only occurs in the South, Essex and the southern West Midlands, contrasting with *litol* which occurs across most of the area covered by the *LALME*. Similarly, while the form *wel* for ‘well’ (item 281) covers almost all of the area covered by the northern *LALME* questionnaire, the form *welle* does not occur in East Anglia and the western half of West Midlands.

<sup>53</sup> E.g. the forms for the nominal plural suffix (item 57), ‘away’ (item 83), ‘enough’ (item 125), and ‘good’ (item 155).

<sup>54</sup> For example the major forms used for ‘together’ (item 268) in this version, *to giddirs* and *togiddirs*, are not attested in *LALME*, and even similar forms (e.g. *to giddir*, *togiders*, *togidder*) are very rare, occurring in only a handful of LPs. Similarly, the forms *tile*, *tyle* and *tyll* which are quite frequently used for ‘until’ (although they are regularised as ‘till’ in the present edition) do not occur at all, while the form *ale* for ‘all’ (item 75) occurs in only seven LPs (the majority of which are localised to northwestern East Anglia and northern Lincolnshire, with single occurrences in Cumberland and Shropshire). Some forms, while occurring in the same area, do exhibit differences in their frequency, like the *k*-forms for cast (item 106), which are considerably less-attested in *LALME* than *cast*.

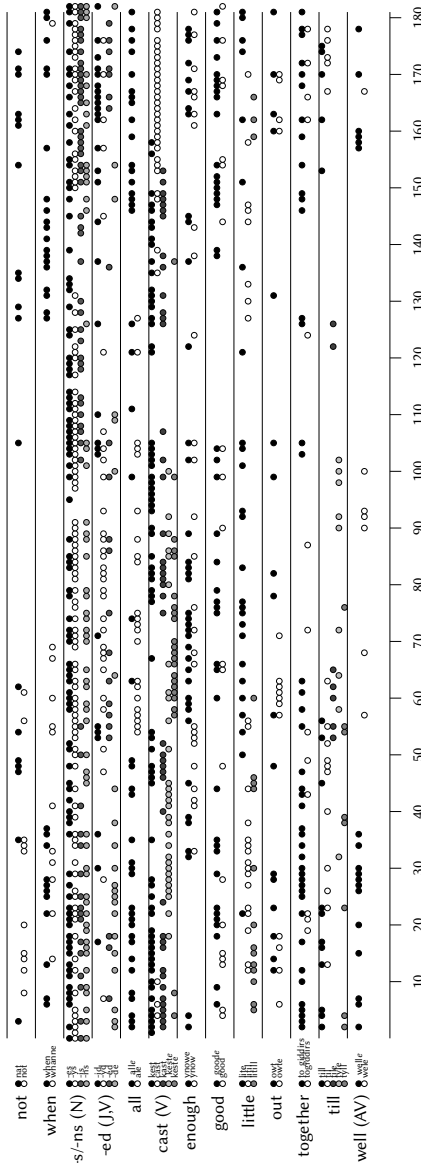


Figure 12.19: The distribution of parallel forms for selected LALME items in terms of the recipes included in MS Additional 5467.

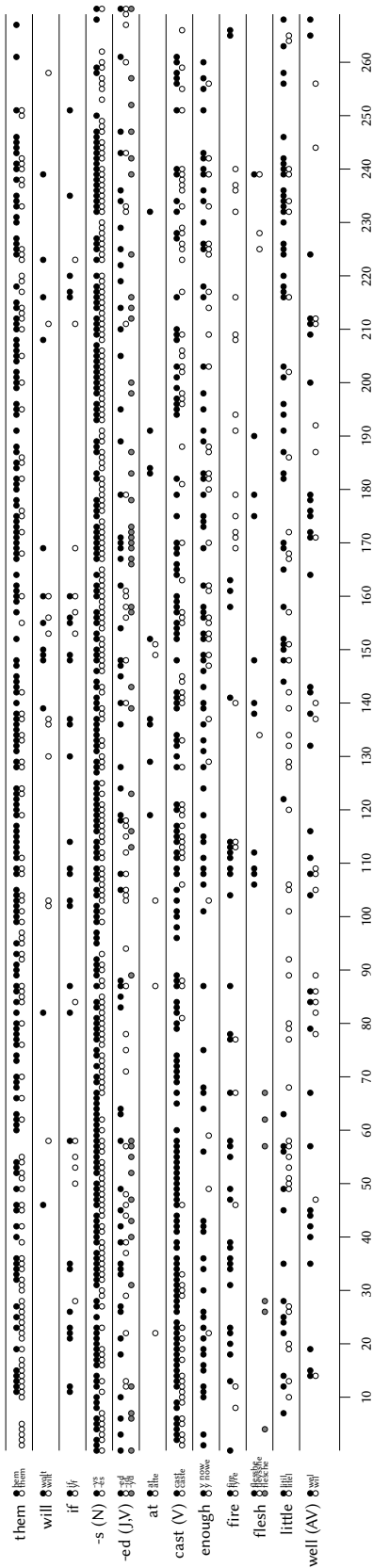


Figure 12.20: The distribution of parallel forms for selected LALME items in terms of the recipes included in MS Ashmole 1439.

series of recipes, possibly indicating the switch of exemplar around this point. The other clearly distinguishable pattern observable in *Figure 12.19* is the increased or decreased occurrence of certain forms in the middle part of the collection. The clearest example of this is provided by ‘well’ (item 281), for which the form *welle* is used in recipes 2–36 and 157–178 and the form *wel* only in the intervening recipes 57–100 (apart from a single occurrence in recipe 167). A similar pattern is also discernible in the forms of ‘all’ (item 75) and ‘little’ (item 191), as well as the forms *til/till* and *tile/tyle* for ‘until’ (item 277).<sup>55</sup> Taken together, these distribution patterns would seem to indicate the use of (at least) two different sources for the main part of the collection (recipes 1–91), the second of which would also seem to be the source of the first subcollection, titled “La manere pur roster *et* saucer diuerse viaundes *et cetera*” (recipes 106–125), and would seem to prefer single ⟨l⟩ over the double ⟨ll⟩, possibly representing a more East Anglian dialect than the sources for the initial and final parts of the collection.

### MS Ashmole 1439

For MS As, the items and forms fulfilling the criteria outlined above all occur widely in the general Midlands–East Anglia area, with differences mainly in their northern and southern extents.<sup>56</sup> As a rather curious detail, there seem to be two forms, *wilt* for the second person singular of ‘will’ (item 24-20) and *yf* for ‘if’ (item 33), that do not seem to occur in the Northamptonshire–Huntingdonshire–Bedfordshire area even though they do occur in the surrounding regions. However, most of the forms examined here are either relatively neutral or even completely interchangeable, judging from their lack of dialectal differentiation and uniform usage pattern in *Figure 12.20*.<sup>57</sup> Some forms—like *atte* for ‘at’ (item 82) and *fleysshe* for ‘flesh’ (item 140) occur so rarely in *LALME* that they cannot be associated with a specific area with any certainty.<sup>58</sup> The distribution of these parallel forms within the collection, however, seems to be less patterned than for the other versions examined so far, and what patterns are discernible do not seem to coincide with each other.

<sup>55</sup> For the first, the form *ale* occurs only in recipes 54–127, while its parallel form *alle* is significantly less frequent in this part of the collection than elsewhere. For the second, the form *lite* occurs significantly more frequently between recipes 50 and 105, and is the exclusive form between recipes 63 and 121. For the third item, the form *til/till* occurs only in the beginning (recipes 2–63) and end (recipes 153–178) of the collection, while the form *tile/tyle* is the predominant form in between, occurring only once (recipe 32) outside of the span between recipes 53 and 126.

<sup>56</sup> For exaple the form *pem* for them (item 8) excludes the North and the very South, while the form *flesche* for ‘flesh’ (item 140), is a relatively northern one, occurring in the North, the Midlands and the north part of East Anglia, but rarely in the South (in contrast to *flesshe*, which is more common in the South and less so in East Midlands). In contrast, both of the forms for ‘will’ (2. sg, item 24-20), *wolt* and *wilt* are limited to the south, East Anglia and southern West Midlands, with *wilt* being more common in the east.

<sup>57</sup> In the former category, one can count the forms for ‘fire’ (item 137) and ‘little’, which occur in most of the area covered by *LALME*, apart from the very north, having only slight differences in their geographic occurrence patterns, and in the latter, the forms for the nominal plural suffix (item 57), the weak past participle suffix (item 65), ‘cast’ (106), ‘enough’ (item 125), and the adverb ‘well’ (item 281) which exhibit neither dialectal differentiation nor distinctive usage patterns within the collection.

<sup>58</sup> *atte* occurs in only four LPs in *LALME*, twice in the North, once in the East Midlands and once in East Anglia, while the only attestation of *fleysshe* in *LALME* is for MS H279.

Of the patterns that are discernible, the ones exhibited by ‘at’ (item 82) and ‘flesh’ (item 140) are of the binary type with a shift between the initial and final parts of the collection, both with a relatively indeterminate *locus*. The shift from *atte* to *at* is initiated at recipe 119, where the latter is first attested, and finishes at recipe 151, which sees the last occurrence of *atte*, the two forms both occurring in between these points, although never in the same recipe.<sup>59</sup> For ‘flesh’, the indeterminacy of the point at which *flesche* gives way to the forms *flesshe* and *fleysshe* is caused by the fact that none of these forms occur in the recipes between 67—the last occurrence of *flesche*—and 106—the first occurrence of *flesshe*. The rest of the macro-level patterns apparent in Figure 12.20 involve the total or near-total disappearance of a form and its subsequent re-appearance later in the collection. These patterns include the absence of the form *yf* for ‘if’ in recipes 84–153, where *if* occurs exclusively; the disappearance of the form *fire* after recipe 163 until its brief reappearance at the very end of the collection (in recipes 265–6), during which the form *fyre* is much more common than in the first half of the collection; and the absence—except for one occurrence in recipe 122—of the form *litol* for ‘little’ from recipes 64–143. While these patterns do not coincide—either with each other or the textual structure of the recipe collection—to the same degree as the ones found in the other versions analysed above, they could be seen to hint at the initial, medial, and final parts of the collection differing somewhat in their dialectal origins. While the boundaries between these three parts are difficult to discern with any certainty—and may in fact be merely coincidental—the patterns described above would seem to place them somewhere near recipe 60 and recipe 120. Since they do not coincide with any discernible shifts in the content of the recipes or any explicit textual divisions, this conclusion should be considered extremely tentative.

### MS Douce 55

For MS D, many of the items and forms fulfilling the formal criteria exhibit some, and occasionally quite significant, degree of dialectal differentiation. For example the three forms used for ‘not’ (item 46)—*nat*, *natt* and *nott*—are quite distinctive in their dialectal distribution, with *natt* being very rare and occurring only three times in Norfolk and once in Warwickshire, *nat* occurring only in the South (apart from the West Country) and East Anglia, and only *nott* occurring sporadically all over the country.<sup>60</sup> This also means that many of the pairs of parallel forms consist of a more neutral form found all across the country, and a more dialectally restricted one.<sup>61</sup> As for the other versions, there are also items—‘away’ (item 83), ‘enough’ (item 125), ‘him’ (item 171), ‘his’ (item 172) and ‘two’ (item 275)—whose forms are clearly not distinctive, as they occur in exactly the same areas in *LALME* and do not exhibit significant patterning within the collection and thus do not

<sup>59</sup> The scarcity of evidence for this item makes it difficult to draw any firm conclusions about this particular pattern, however.

<sup>60</sup> An extreme example is ‘if’ (item 33) for which both of the major forms, *yiffe* and *yiff*, are very rare in *LALME*, the former occurring only once (in Surrey) and the latter six times (four of which are in East Anglia, one in Surrey and one in northern Gloucestershire).

<sup>61</sup> For example the forms *gode* (item 155), *other* (item 222), *out* and *oute* (item 225) occur in all areas covered by their respective questionnaires, while the forms *good*, *oper*, and *outt* have more restricted areas of occurrence, being limited to the southern Midlands and East Anglia (*good* and *oper*) or to northern Lincolnshire and Yorkshire (*outt*).

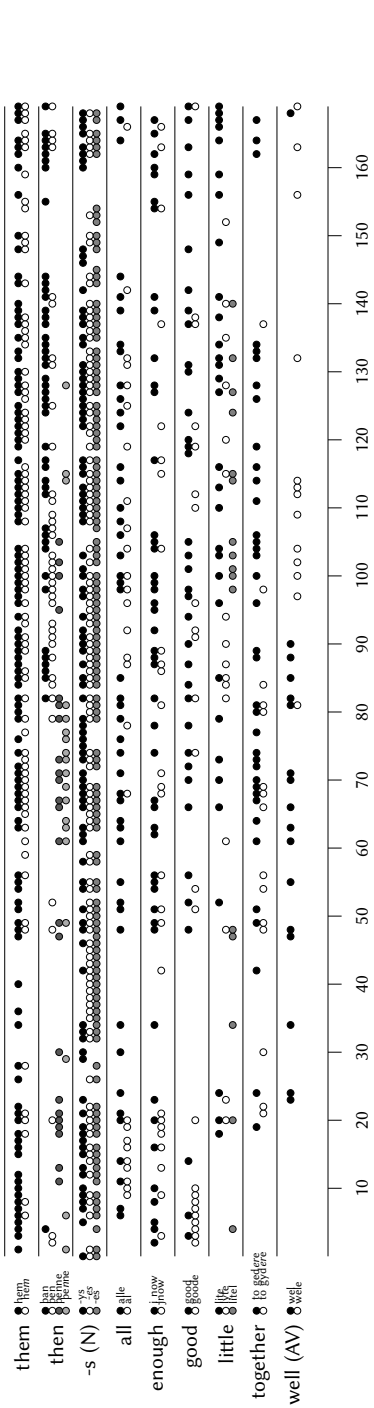


Figure 12.21: The distribution of parallel forms for selected LALME items in terms of the recipes included in MS Cosin V.iii.11.

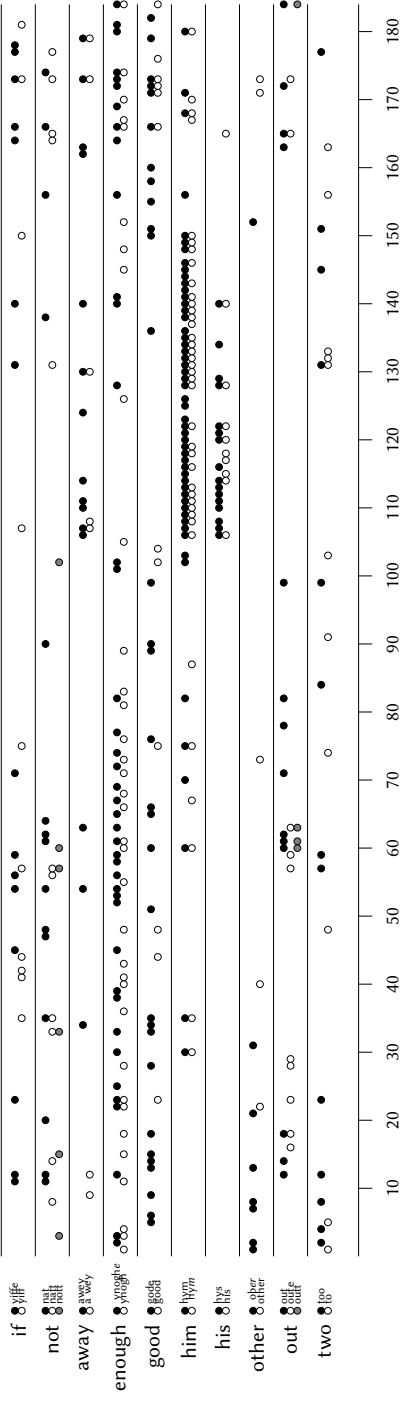


Figure 12.22: The distribution of parallel forms for selected LALME items in terms of the recipes included in MS Douce 55.

provide useful evidence for dialectal variation within the collection.<sup>62</sup> Despite being dialectally differentiated, few of the forms listed in *Figure 12.22* exhibit any significant large-scale patterning within the collection.

Of the scant patterns that can be distinguished, perhaps the most noticeable is the fact that despite being the most common form of the three in *LALME*, the form *nott* is not only the least-attested one in this text but also occurs only in the first half of the collection up to recipe 102. Also the typically Midlands form *oper* for ‘other’ occurs almost exclusively in the beginning of the collection (up to recipe 31) but only once in the rest of the collection (in recipe 152), *other* being used instead.<sup>63</sup> The distinctive form *outt* for ‘out’ occurs only in three recipes grouped very closely together (recipes 60, 61 and 63), which also roughly coincide with occurrences of the form *nott* (and *natt*) for ‘not’, possibly indicating a distinct source for the sequence of recipes around this point, preferring the dual-t forms.<sup>64</sup> While there are also some sections of the collection where certain forms are absent—such as the stretch between recipes 12 and 107 for *a wey* and the stretch between recipes 63 and 165 for *oute*—the low number of occurrences for the corresponding parallel forms in these parts of the collection means that these absences could be merely coincidental. Thus, the mixture of forms in the MS D version of *PD* would not seem to result from any obvious combination of dialectally disparate exemplars, but is more likely to reflect either a single, relatively heterogeneous dialect with several native forms for each item, or the result of multiple rounds of inconsistent copying resulting in the dilution of any boundaries between sections with different dialectal origins. The fact that most of the parallel forms do at some point occur within the same recipe would seem to support the first hypothesis.

### MS Cosin V.iii.11 A

For MS C, most of the forms fulfilling the criteria for this analysis would seem to exhibit only slight dialectal contrast between them, although for several of the items, all of the forms are strongly associated with a specific dialect area, the most notable of these being the forms for ‘enough’ (item 125) and ‘together’ (item 268).<sup>65</sup> Most of the forms that exhibit some dialectal contrast among themselves only differ in the extent to which they are found in the far north or south, all occurring

<sup>62</sup> However, as can be seen in *Figure 12.22*, they can still provide other kinds of information about internal variation; for example the forms for ‘him’ (item 171) and ‘his’ (item 172) indicate an interesting stylistic feature exhibited by the section containing recipes for fowl and fish in this version of the collection, namely that it consistently uses personal pronouns to refer to these animals, making them overwhelmingly more frequent in this section of the collection than elsewhere.

<sup>63</sup> Although the low number of occurrences for this item makes it difficult to draw any significant conclusions from this pattern.

<sup>64</sup> Although the fact that the recipes around this point do not seem to constitute a coherent unit either in terms of their content or their order of occurrence in the other versions (see chapter 13).

<sup>65</sup> The forms *jnow* and *j now* principally used for ‘enough’ (item 125) in this version are as such not really attested in *LALME* (the form *j now* being recorded twice for East Anglia and *jnow* not at all), but assuming that they are equivalent to the forms *I now* and *Inow* (the tall ⟨i⟩ being used as a capital initial), they are both limited to East Anglia and the coastal area slightly to the north, although they do not seem to be distinctive in relation to each other. The specific forms recorded here for ‘together’ (item 268) also occur in the same limited area in East Anglia, although similar unabbreviated versions also occur in a slightly wider area extending to the South and the southern Midlands.



equally in the middle area of the country.<sup>66</sup> As in the other versions, there are also some frequently occurring parallel forms that exhibit neither dialectal nor distributional differences. These include the forms for the nominal plural suffix (item 57, most likely governed primarily by the headword) and the abbreviated and non-abbreviated variants for 'them' (item 8).

In addition to being dialectally relatively uniform, the parallel forms analysed here are also relatively undifferentiated in terms of their spatial distribution. The only clear binary distribution in this version occurs with the adverb 'well', for which the initial part of the collection, up to recipe 90, uses the form *wel* while the final part exclusively employs the form *wel* apart from a single occurrence of *wel* in recipe 168 at the very end of the collection.<sup>67</sup> A slightly less total but nevertheless clearly discernible shift, located at roughly the same point, occurs in the forms for 'then' (item 30), with the initial part of the collection up to around recipe 80 strongly favouring the form *þenne/þenne*, after which this form becomes much rarer and the forms *þan* and *þen* become dominant. All of the other discernible patterns involve either the absence or increased presence of a specific form in a specific section of the collection. These include the absence of the form *al* for 'all' (item 75) in recipes 21–67, the absence of the form *litel* for 'little' (item 191) from recipes 49–97 (and from recipe 140 onwards), and the disappearance—apart from a single occurrence—of the form *to gydere* for 'together' (item 268) after recipe 98. Taken together, these patterns would seem to suggest that the initial and final parts of the collection, the dividing line being somewhere around recipe 80,<sup>68</sup> being copied—either by the scribe of this MS or by one of his predecessors—from two different sources with slightly different dialectal profiles,<sup>69</sup> although the other, non-coinciding patterns described above make the picture somewhat unclear.

## Conclusions

The considerable variation in the ordering of the recipes in the six collections—discussed below in chapter 13—combined with the difficulty of determining the exact location of the possible dialectal breaks identified here makes it difficult to compare the degree to which they coincide in the different versions, the only case readily apparent from the analyses here being the close coincidence of the dialectal shift observed roughly halfway through both MS C and MS H4016, possibly

<sup>66</sup> For example the form *alle* for 'all' (item 75) does not really occur north of Durham in the northern LALME questionnaire, while the form *wel* for 'well' (item 281) seems to be more common in the north than the alternative *wel*, but in the Midlands or East Anglia there is no difference between the two forms of either item. Similarly, the form *goode* for 'good' (item 155) is slightly less common in the very south than its parallel, but both occur equally elsewhere in the country. Perhaps the most significant of these differences are found between the forms for 'little' (item 191), of which *lite* and *lyte* are limited to the South and East Anglia, while *litel* occurs also in the Midlands and the North, and between those of 'then' (item 30), of which *þen* would seem to be slightly more northern than *þan*, while *þenne/þenne* are characteristic to the West Midlands.

<sup>67</sup> The fact that *wel* first co-occurs with *wel* in recipe 81 before becoming dominant might be the result of the scribe first encountering it at that point, either due to changing to a different exemplar or due to a dialect shift in an exemplar itself, but continuing to still use *wel* for some time before adopting the spelling of the exemplar.

<sup>68</sup> This shift in orthographic forms does not seem to coincide with a content-based or text-structural division in the collection, which means that it remains largely hypothetical.

<sup>69</sup> The specific characteristics of these profiles would require further study, possibly through the preparation of separate profiles for the two parts.

pointing towards a change of exemplars at this point in a shared ancestor.<sup>70</sup> Nevertheless, the different kinds of patterns described above—occasionally coinciding with clear structural divisions that would make for natural boundaries between recipes copied from different sources—make it likely that most of the six versions of *PD* are not only dialectally different from each other, but reflect diverse origins also internally and contain sections of recipes from dialectally diverse sources. It should be noted that this does not mean that versions—or individual sections of recipes—with differing dialectal features would necessarily have different ultimate origins, but may merely have been undergone a different set of dialectal translations during their textual transmission, although it is likely that at least part of the variation in the content of the six versions is caused by the addition of new recipes from originally disparate sources.

The differences in not only the order but also the number of recipes between the six versions also mean that the macro-level patterns discussed above are also likely to be matched by (potentially more significant) micro-level patterns resulting from the copying of short segments of recipes (or even individual recipes) from different exemplars. Since these kinds of smaller-scale patterns are difficult to detect by the exploratory methods employed here, a more detailed study—based on the generation and analysis of detailed linguistic profiles both for each version of each individual recipe and for the groups of recipes identified in chapter 13 on text-structural grounds—will most likely provide more information both about the specific textual history of this particular family of recipe collections, and more generally about the ways in which the individual components of discourse colonies are transmitted in a manuscript environment.

### 12.4.2 Relationship to Chancery English and the emerging standard

While the linguistic profiles of the six versions of *Potage Dyvers* exhibit considerable variation on the level of individual forms, as indicated by the relatively modest degree of agreement exhibited in *Table 12.8*, they also share a large number of forms, many of which imply a predominantly London, Essex or a West Midlands origin, as was shown above in section 12.3. This kind of preponderance of Midlands and Essex features in a late Middle English text does not necessarily indicate a geographical origin in these areas. As was explained in chapter 6, the dialect of London—which had originally been a southern one—had by the 15<sup>th</sup> century been influenced strongly by the East Midland dialect and agreed with it in most respects. Being the language of the capital, this was also the dialectal variant that most influenced the language of other regions, and by the latter half of the 15<sup>th</sup> century, it had been accepted as a standard in most parts of the country, making it hard to identify the region of origin of a text on linguistic grounds. After the first quarter of the century (to which period the earliest of the manuscripts under study have been dated), there is already found an increasing number of manuscripts that have escaped localization by the compilers of *LALME* because of their advanced level of standardisation (Taavitsainen 2004: 209).

<sup>70</sup> These versions which share a large proportion of the same recipes and are organised quite similarly with only minor variation in the order of the recipes (see chapter 13).

Because of the relatively late date of the *PD* manuscripts and the prominence of dialectal features occurring in the London region and western Essex in the *PD* manuscripts, it seems prudent to also examine their relationship to the specific language variety that was adopted first by Henry V's private secretariat and subsequently by the Royal Chancery, whose voluminous correspondence and training of scribes is believed to have spread it around the land. In terms of its orthography, Fisher (1977: 884) characterises Chancery English as

- 1) preferring the ⟨gh⟩ spelling of the native palatals (e.g. in 'high' and 'through'),
- 2) eschewing the northern forms *wich/wech* for 'which',
- 3) not exhibiting word-initial ⟨q⟩ in words beginning with ⟨wh-⟩,
- 4) continuing to use ⟨o⟩ before ⟨n⟩ in words like *lond* and *stond*, and
- 5) using the characteristically Chancery spellings of *eny* for 'any' and *shew* for 'show'.

In terms of the first feature, the forms used for the words 'high' and 'through', used as examples by Fisher (1977: 884), were examined in each of the MS versions of *PD*. While none of the instances of 'high'—numbering between zero and four—contain ⟨gh⟩,<sup>71</sup> its occurrence in the forms of 'through' varies significantly from one manuscript to the next: while MSS Ad and C contain no forms with ⟨gh⟩, all of the forms found in MS H4016 end in ⟨gh⟩.<sup>72</sup> For the second feature, the northern forms of 'which' do not occur in any of the versions. For the third feature, the only instance of initial ⟨q⟩ with words whose modern forms begin with ⟨wh-⟩ is the form *quyte* for 'white' in MS Ad, the other versions containing none. For the fourth feature, occurring in the *PD* in the words 'hand' and 'stand', MS Ad uses both of the forms to equal degree (17 instances of ⟨-and-⟩ and 18 of ⟨-ond-⟩), while the other versions are heavily biased towards the ⟨-ond-⟩ form, the most extreme bias being found in MS H4016, which contains 34 forms with ⟨-ond-⟩ and none with ⟨-and-⟩.<sup>73</sup> As to the fifth feature, the latter word does not occur in the recipes, but the former occurs in all of the manuscripts except for MS C, exhibiting the form *eny* in MS Ad, MS H4016 and MS D, it being the exclusive form in the latter two (with 2 and 4 occurrences, respectively) and occurring once (with two occurrences of *any*) in MS Ad.

However, as Fisher points out, the standardisation of Chancery English was much more marked in its morphology than its orthography, exhibiting the following features:

- 1) second person singular pronoun is always *ye/you*;
- 2) third person plural pronoun is quite regularly *they, them, their*, although *hem* and *her* also occur sporadically;
- 3) the reflexive *-self/-selves* is frequent;
- 4) adverbs never end in *-lich*;
- 5) verbs rarely have the plural *-n*;
- 6) the participial *-n* is found on many words which have since lost it, but the

<sup>71</sup> The forms used by the manuscripts include *hy* (MSS Ad, C and H279), *hye* (MS As), *hi* (MS C), and *hie* (MS H4016).

<sup>72</sup> The other manuscript versions lie between these extremes, MS H279 exhibiting only a single example out of 100, while MS As exhibits it for three fourths of the total (77/105) and MS D for almost all forms (48/51).

<sup>73</sup> The rest of the versions exhibit the ⟨-ond-⟩ and ⟨-and-⟩ forms in the following numbers: MS As, 21/10; MS D, 25/5; MS C, 28/2; and MS H279, 30/2.

- participial *y-* has been completely lost;
- 7) the preterite form is always with ⟨-d⟩ and never with the northern ⟨-t⟩;
  - 8) the negative particle is placed after the verb (“be noght able” vs. “ne be able”);
  - 9) the third person singular suffix of verbs continues to be ⟨-th⟩; and
  - 10) the continued use of *be/ben* over the eventually adopted northern *are*.

Also these morphological features are attested to varying degrees by the six manuscript versions, although they betray much less similarity with Fisher’s description of Chancery English. With regard to the first feature, none of the versions resemble Chancery English, using the ‘thou/thee/thy/thine’ forms in overwhelming quantities in relation to the ‘you’ forms which occur very rarely.<sup>74</sup> For the second feature, the majority of the versions—with the notable exception of MS As—favour the ⟨h-⟩ initial forms, the Chancery English forms being in the clear minority.<sup>75</sup> The exception to this is MS As, which clearly favours the ⟨th-⟩ forms (including those beginning with ⟨þ⟩), which account for 92 per cent of all forms for the 3<sup>rd</sup> person plural pronouns. Although Fisher’s characterisation of the reflexive forms as “frequent” in the third feature is a subjective one, it does not seem to be attested in any of the *PD* versions, as the largest number of reflexive forms is 11 (MS H279), and most of the versions contain only one or two instances and MS C contains none. In terms of the fourth feature, the versions are divided in two—while MSS As, D and H4016 resemble Chancery English in not containing adverbs ending in *-lich*, MSS Ad, C and H279 all contain at least one example of such forms. For the fifth feature, MS Ad is the only version that does not use the plural ⟨-n⟩ at all, the others using it at least occasionally on the verb ‘be’, MSS As, MS C and MS H4016 using it relatively consistently, and MS H279 using it not only on ‘be’ but also on a variety of other verbs.<sup>76</sup>

In terms of the sixth feature listed by Fisher, all of the versions except for MS D have past participles with the ⟨y-⟩ or ⟨i-⟩ prefix, although they also do contain several strong forms with the ⟨-n⟩ suffix. For the preterite form (feature 7), no forms ending in ⟨-t⟩ are found in any of the versions, but on the other hand very few preterite forms of any kind occur in the recipes. The negative particle (feature 8) appears both before and after the verb in all of the *PD* versions, being about twice as common in the position following the verb. The third person singular suffix is exclusively ⟨-th⟩ in all other versions but MS D, which also contains a single example of the ⟨-s⟩ suffix (*seethes*). In terms of the last feature in the list, all of the *PD* versions conform to the Chancery English model in that the form *are* does not occur in any of them.<sup>77</sup>

If we look at the summary of these features for each *PD* version in *Table 12.10*, we can see that the different versions exhibit Chancery forms to different degrees. MSS H279, Ad and C—the first of which is localised to the southwestern West Midlands, the second either to the same area, the Southwest or to the area around

<sup>74</sup> The proportions of the *thou* and *you* forms in the different versions, from lowest to highest, are 76/7 in MS Ad, 103/9 in MS H4016, 73/5 in MS D, 62/3 in MS C, 226/1 in MS As, and 269/1 in MS H279.

<sup>75</sup> The ⟨th-⟩ forms account for only 3–6 per cent of the forms in all MSS except for MS As.

<sup>76</sup> Verbs that occur with a plural ⟨-n⟩ in MS H279 include ‘do’, ‘break’ (*tylle þey brekyn*), ‘say’ (*men sayn it is*), and ‘come’ (*cruddys þat comen*).

<sup>77</sup> Although the frequency of subjunctive forms in the recipe text type tends to create ambiguity in some cases as to whether a specific verb is in the indicative or subjunctive mood.

MS version	1. <gh> spelling	2. Absence of <i>wich/ wech</i>	3. No <q> for <wh>	4. <o> before <n>	5. <i>eny</i> for 'any'	1. Exclusive 'you'	2. Preference of <i>they</i> forms	3. Frequent reflexive <i>-self</i>	4. No <i>-lich</i> adverbs	5. Verb plural <-n> rare	6. Loss of participial <y->	7. Preterite <-d> instead of <-t>	8. Negative after verb	9. 3 p.s. suffix <-th>	10. Use of <i>be/ben</i> for 'are'
Harley 4016	+	+	+	+	+	-	-	-	+	-	-	+	-	+	+
Harley 279	-	+	+	+	-	-	-	-	-	-	-	+	-	+	+
Additional 5467	-	+	-	+	+	-	-	-	-	+	-	+	-	+	+
Ashmole 1439	+	+	+	+	-	-	+	-	+	-	-	+	-	+	+
Douce 55	+	+	+	+	+	-	-	-	+	+	+	+	-	+	+
Cosin V.iii.11 A	-	+	+	+	-	-	-	-	-	-	-	+	-	+	+

Table 12.10: A summary of the Chancery English features (Fisher 1977: 884) in the PD versions.

London in western Essex, and the third to the border of Essex and Suffolk—share only few of the features described by Fisher (1977) as typical to Chancery English and are unlikely to have any significant connection to this specific language variant. On the other hand, MS D exhibits most of the features to the extent that it might well be written by a scribe used to reading and writing Chancery English, either in London or—considering the mid-15<sup>th</sup>-century date of the text—in some other area influenced by the Chancery standard, although the *LALME* analysis in subsection 12.2.5 would seem to support a London origin. The rest of the versions, MSS H4016 and As do exhibit more than half of the Chancery features but also features that are not usually exhibited by Chancery English, such as the participial prefix <y-> and the verb plural suffix <-n>. These texts, representing a southern West or Central Midlands and western Essex dialects, respectively, are unlikely to represent direct influence of the administrative standard, but the shared features are more likely the result of the general proliferation and spread of these features in the mid- and late-15th century.

12.5 Conclusion

The conclusions to be drawn from this exploratory analysis of the dialectal nature of the *Potage Dyvers* manuscripts—much like this thesis in general—relate equally to the methodology used as to the actual object of inquiry. In terms of the methodology, the analysis process undertaken here proved that the regularisation and part-of-speech annotation of the text in the digital edition allows for the extremely efficient automatic creation of relatively detailed and accurate linguistic profiles through the application of scripts that operationalise the *LALME* questionnaire into a set of search criteria. Not all of the items of the *LALME* questionnaire can be operationalised as easily, however. While questionnaire items based on lexical items can be targeted explicitly and unambiguously and thus automatically quantified, the restriction of grammatical annotation to basic word-class tagging means that many grammatical forms need to be retrieved indirectly based

on a combination of the word-class information and regular expressions targeting components of the regularised and original forms, introducing ambiguity and requiring manual pruning of the search results, and thus precluding the automatic calculation of occurrences.<sup>78</sup>

While a digital edition containing the full text of a document allows dialectal forms and other linguistic phenomena to be quantified in absolute terms, the fact that *LALME* itself is based on an ‘analog’ analysis of partial samples—and thus does not provide absolute quantities—precludes the use of these absolute quantities as such. Fortunately *LALME* explicitly describes the mathematical relations underlying its categorisation of forms according to their relative frequencies, which allowed the generation of *LALME*-compatible linguistic profiles from the absolute quantities recorded from the digital edition. The use of comprehensive, detailed linguistic profiles for localising manuscripts using the fit-technique also revealed some problems inherent in *LALME* itself, most of which derive either from the manual compilation process or the separation between the northern and southern parts. Perhaps the most significant of these is the apparent inconsistency in the recording of some items, such as the present forms of the verb ‘have’ in the southern part, for which only uncommon or otherwise noteworthy forms seem to have been recorded, judging from the drastic difference in its frequency between the northern and southern parts of *LALME*, which is unlikely to reflect actual occurrence patterns. Another type of inconsistency is caused by the existence of several methods for recording some items, which makes systematic analysis difficult. For example the southern past participle prefix is in some LPs recorded as the specific form used, whereas others merely record its presence using the value X.<sup>79</sup> Since the ‘fit’ technique is based on the gradual exclusion of those areas which do not exhibit the forms used in the text to be analysed, LPs that do not contain any value for a given questionnaire item are problematic, since there is no data on which to base their exclusion or inclusion (*LALME*: vol. 3, xiii). As described above in section 12.1, the method adopted here for mitigating the effects of this and minimising the chance of excluding areas merely due to sampling anomalies was to extend the effect of individual matching LPs relatively far into the surrounding area in order to compensate for the high likelihood of only some LPs within a dialect area containing an example of any given item of the questionnaire.

In the above analysis, the fragility of the *LALME* profiles and the fit technique with regard to the accidental omission of individual forms or items is demonstrated by the differing localisation of MS H279 reached here and established in *LALME*. Partially this difference is explained by the fact that although it is prepared from an edition of the entire text (Austin 1888), the *LALME* profile for the MS does not record all of the forms actually found in it (see footnote 20 on page 549), most notably omitting the southern past participle prefix ⟨y-⟩, which is in fact

<sup>78</sup> The grammatical features whose annotation would provide the most value in terms of the *LALME* questionnaire would be the mood, tense and person of verbs and the number of nouns.

<sup>79</sup> These problems are apparently a result of the fact that the two parts were compiled separately by two different people following different principles: “SOU [southern] analyses tend to be selective once a stable pattern of usage has been defined. Commonly they are fairly complete records for part of the text, supplemented by scanning for items not found in the initial sample. NOR [northern] analyses, by contrast, tend to be full reports for more strictly delimited samples, and are relatively seldom amplified by scanning beyond.” (*LALME*: vol. 3, xii)

very common in MS H279.<sup>80</sup> However, the main reason for the difference is most likely the selection of different forms as the basis of the analysis, since as *Figure 12.6* shows, none of the surrounding LPs are very similar to MS H279 in terms of the diagnostic forms chosen here, the Herefordshire–Gloucestershire border providing both closer individual matches and a greater level of overall similarity.<sup>81</sup> This dependency on the choice of diagnostic items emphasises the importance of basing the localisation on forms that occur consistently and can thus be considered to be typical to the scribe, instead of rare forms whose status in the scribe's repertoire is less certain.<sup>82</sup>

In terms of the dialectal features of the *Potage Dyvers* versions, the analyses undertaken here would seem to suggest that the textual tradition of the *Potage Dyvers* collection is a broadly southern one. While there are clear differences in the linguistic profiles of all the six versions, resulting in combinations of forms typical to different areas of the south, there is also a relatively large body of shared forms. As was observed above, this shared body of forms consists mostly of widely used forms, most likely indicative of a relatively advanced state of linguistic standardisation rather than of the dialectal features of any shared ancestor. As was hypothesised in subsection 6.2.2, the language forms of late Middle English texts like the *PD* are no longer likely to represent individual dialects, but rather hybrid accommodations to the different idiolects used in the discourse communities in which they circulate, employing a selection of forms that are likely to be understandable and even familiar both to their copier and to their intended audiences. Since all discourse communities would most likely have had multiple forms they found acceptable, i.e. that belonged to their passive repertoire, the process of constrained selection, described by McIntosh, Samuels and Benskin (*LALME*: vol. 1, 18-9), combined with the kind of copying from multiple sources that is typical to discourse colonies like recipe collections, would tend to produce not only linguistic profiles combining forms originating in different dialect areas, but also the kind of subtle and spatially structured variation that was observed in subsection 12.4.1.

On the other hand, the existence of differences in the dialectal profiles of the six versions despite their relatively late date reveals that standardisation was still far from complete, and the majority of English lexis still exhibited several morphological and orthographical variants, even if their occurrence patterns were more often indicative of equivalence and co-occurrence rather than mutually exclusive and dialectally distinctive alternation. Rather than standardisation in the sense of homogenisation, the indiscriminate intermingling of orthographically distinctive parallel forms would seem to reflect the normalisation of heterogeneous idiolects and the weakening of dialectal identity. Instead of translating the orthography of the recipes into a single homogenous linguistic profile, the scribes seem to be content with employing a variety of forms—often with disparate dialectal origin—for the same lexical item, often even within a single recipe.

<sup>80</sup> On the other hand the profile frequently exaggerates the frequency of minor items in relation to the quantificatory guidelines described in McIntosh, Samuels and Benskin (*LALME*: vol.3: xiv).

<sup>81</sup> Here the relatively frequency-agnostic nature of the present methodology makes the match more convincing, as it is not significantly influenced by the much greater overall frequency of LPs in the Herefordshire–Gloucestershire area compared to the Sussex area.

<sup>82</sup> It is unfortunate that *LALME* does not provide information on the specific forms that contributed most significantly to the localisation of each LP.

In terms of the degree of dialectal similarity between the six versions, the total similarity score between two versions seems to be a relatively poor predictor of the geographical distance of the likely areas of origin indicated by the 'fit' method of Benskin (1991) using the most frequent dialectally distinctive forms of each text. For example, the most likely localisations for the pair of versions found to be least similar in terms of their dialectal features, MSS H279 and Ad in fact overlap in the West Midlands, mostly due to the fact that the area of origin for MS Ad is very indeterminate and covers several disparate areas. And while the two versions found to most resemble each other in terms of their similarity score, MSS C and As, are indeed localised in the same area in Essex, MS C also has a very high similarity score in relation to MS H279, which on the other hand would seem to most likely originate near the Welsh border. This lack of correlation between the overall level of agreement in dialectal forms and the likely area of origin again highlights the extremely variable dialectal significance of different forms and the importance of basing any localisations and similarity metrics on forms that are both characteristic of the text and dialectally distinctive. Thus the kind of similarity metric used in subsection 12.3.1 will need to be developed further by adding relative weights to different forms based on their dialectal significance. Since this significance is dependent both on the geographic distribution of the forms and on their frequency in the text, the resulting formula will by necessity be relatively complex and the formulation and testing of suitable candidates will require further research.

While it is difficult to distinguish separately transmitted parts of the recipe collection merely on the basis of linguistic evidence, the analysis of the internal variation of the individual manuscript versions would seem to support the observation made by recipe scholars that as recipe collections get copied, they tend to accumulate and integrate material from a number of sources. While the method employed here, based on the spatial distribution of selected diagnostic features, provides only a partial view of the total internal variation, it does provide a useful starting point for both the further study of this particular family of collections by indicating likely candidates for subdivisions for which to generate and compare independent dialectal profiles, and for the study of discourse colonies in general using these kinds of ordered linguistic profiles. The establishment of several consecutive linguistic profiles for the dialectally discrete component parts of a discourse colony text can not only provide information about the process of its compilation, but will also enable the dialectal similarity and other comparative metrics to be calculated and compared independently for these dialectally more homogenous component parts. This, in turn, could allow dialectal relationships and likely areas of origin to be determined separately for these components, providing more information about the processes of selective copying and compilation of collections from disparate sources.

While all of the six surviving copies of the *PD* show signs of having been copied in different parts of the South and the Midlands—with the exception of D which shows a higher likelihood of originating in London—the textual tradition of the collection as a work is likely to have crossed London at some point, not only because of the linguistic evidence, but also owing to the evidence of many European culinary collections originating in royal households (see subsection 8.4.2), and the observation of Taavitsainen (2004) that “London is the most likely place for specialised book trade, and metropolitan language use may provide an explanation



for unexpected forms” (211).<sup>83</sup> Assuming London as the region of origin would also help explain the prolific co-occurrence in the same text of features typical not only to the environs of London but also to the West Midlands and East Anglia; if we accept the hypothesis that the changes taking place in the London dialect in the early 15<sup>th</sup> century were shaped to a great extent by immigrants not only from the Central Midlands but also from the North and the West Midlands (Taavitsainen 2004: 212), the forms brought together by them in London would be ones also occurring in these areas.

---

<sup>83</sup> Voigts (1989: 384-6) has observed that although the beginnings of professional book production in London have traditionally been associated with the advent of printing, they can in fact be traced back to the earlier part of the 15<sup>th</sup> century and into commercial scribal workshops specialized in secular manuscripts. It is not at all unlikely that a version of the *Potage Dyvers* would have been commissioned or bought from such a workshop, especially considering the proficient character of the hand used for copying most of the versions.



## Chapter 13

# Structural relationships

The organisation and structure of any text is not only an important but also highly prominent part of its identity. Since “handwritten books are always involved in complex mediations, not simply with their local conditions of production, but also with their pasts, with predecessor volumes” (Hanna 2000: 99), the textual organisation—just as the language and visual layout—of a manuscript copy of a text is always influenced by that of its textual tradition, manifested to the copyist in the form of the exemplar. This means that the analysis of the varying textual structure in different versions of a text can provide interesting information about the transmissional history of the text. This is all the more true in the macro-level organisation of *discourse colonies* like recipe collections, which is not intrinsic to the meaning and identity of the colony but can be freely accommodated to new situational and cultural contexts.

The fact that the organisation of the structural components of a discourse colony does not influence its meaning has two important implications for the copyist. First of all, the range of potential variation in text structure is much greater than in conventional texts, as virtually any permutation of the component parts is not only possible but also potentially appropriate. For example in the case of recipe collections, recipes can be organised not only on the basis of their intrinsic features such as the main ingredient, the type of dish described, the principal method of cooking employed, the time of the year the dish is intended for or the customary position of the dish in a meal, but also on such extrinsic features such as their popularity with a patron, their order of acquisition, the source from which they were copied, or, indeed, the order in which they were presented in the exemplar. In addition to the large number of possible paradigms, the number of potential textual configurations is further increased by variance in the interpretation of these paradigms and their relations to individual recipes—as was noted in subsection 7.1.2, the observation of Lent and other days of abstinence, and consequently also the status of specific dishes in relation to their restrictions, was interpreted differently by different people in different times, and the classification of foodstuffs in terms of their nutritional or humoral properties was far from standardised.

Unlike changes in dialectal or orthographic features, which are more subtle and likely to be the result of a complex interplay between the linguistic resources

of the copyist and the stimulus provided by the exemplar, changes made—or left unmade—by a copyist to the textual structure of a copy are likely to be the result of conscious and deliberate choices. While changes in the dialect or orthography of a manuscript copy can come about as the result of the scribe merely not paying enough attention to the language of the original and copying the work in his own native language variety, significant changes in the textual organisation of a recipe collection beyond the accidental omission or transposition of the odd recipe require a conscious effort, and possibly even the use of mechanical aids such as the marking off the recipes already copied in the exemplar.<sup>1</sup> This deliberate nature of text-structural alteration means that unlike linguistic change, which is often gradual, partial and subtle, changes in textual organisation are often the result of deliberate redesign resulting in drastic and noticeable changes occurring at a single transmissional point. From the point of view of textual transmission, this means that they are likely to produce strongly bifurcating textual histories, as the textual organisation is likely to be either preserved (more or less) as it was in the exemplar or changed in a significant and clearly discernible way.

In addition to being more likely to be deliberate, changes in the textual structure of a text are also more likely to be *functional* than changes in its language. While the translation of a text into the native dialect of its copyist or the modernisation of the language of an older text is functional in the sense that it makes the text more readily understandable to its intended audience, it does not necessarily reflect a change in the function of the text itself, but rather an accommodation to a new situational and cultural context. On the other hand, a change in the textual organisation of a discourse colony like a recipe collection, where the organisation of the component texts does not affect their meaning, can be interpreted as a reflection of either a change in the access strategy employed in using the text or at least a different view of the effectiveness of a specific access strategy in relation to the intended function of the text. Thus the analysis of the textual organisation of a discourse colony in terms of the semantic content of its components parts—e.g. the types of dishes described by recipes—can provide us with information about the conceptual categories and classifications current in its original cultural context. It could also be argued that for discourse colonies with their potential for division and recombination, the analysis of the macro-level organisation of the textual components within the colony is—or should be—a necessary precondition for any study of the internal micro-level organisation of those components themselves, as it can alert us to possible divisions and disjunctions between parts of the colony with differing textual histories and avoid the conflation of linguistic or textual features that originate in potentially different linguistic, cultural and situational contexts. This chapter will examine the text-structural organisation of the six versions of the *PD* family in order to both establish their relationships to each others in terms of their textual ancestry, and to detect any groups of recipes that occur as consecutive sequences in multiple versions and are this likely to have been perceived as established thematic groups.

<sup>1</sup> While it is of course impossible to know for certain the purpose of the various kinds of annotations added to MSS, the Ad, C and H4016 versions of *Potage Dyvers* contain marginal markers whose function might have been exactly this.

## 13.1 Structural similarity of the versions

The structural relationships between the different versions of the *Potage Dyvers* as separate but related discourse colonies are here approached not only in terms of the number of recipes shared by them, but also of their ordering within the collection. In order to analyse the degree of organisational similarity between the different versions of *PD*, both of these aspects need to be quantified in some way. The natural starting point for this quantification is the number of recipes that occur as *parallel versions* in two or more versions, in the sense defined in subsection 10.3.2. The number of shared material, however, is only one dimension of the structural similarity, another being the *organisation* of these shared recipes. In the present analysis, structural similarity is seen as composed of these two factors: similarity of *extent*, i.e. the proportion of recipes shared by the versions being compared, and similarity of *order*, i.e. the proportion of these shared recipes that occur in the same sequence in the compared versions.<sup>2</sup>

Of these two measures of similarity, the quantification of the similarity of *extent* between a set of versions is relatively trivial and can be done similarly to the calculation of shared dialectal forms in subsection 12.3.1 (although without the need to account for different levels of relative frequency), using the following simple formula:

$$\frac{n \times s}{t_1 + \dots + t_n}$$

where

$n$	number of versions being compared
$s$	number of recipes shared by all $n$ versions
$t_x$	total number of recipes in version $x$ being compared

Quantifying the similarity of the *order* of the recipes within different versions of the collection is much less trivial, as there is no obvious or established measure for it. Of traditional sequence similarity metrics, the one best suited for the present purpose is the *Levenshtein distance* (Levenshtein 1966) or its extended version, the *Damerau–Levenshtein distance* (Damerau 1964).<sup>3</sup> These metrics measure the *edit distance* between two sequences—traditionally character strings—i.e. the number of edit operations required to produce one string from the other, the edit operations being *deletion*, *addition* and *replacement* of items, to which the Damerau–Levenshtein distance adds the *transposition* of two adjacent items as a single operation.<sup>4</sup> These metrics, which have been developed for comparing words and other character strings, give more weight to the absolute positions of items in the

<sup>2</sup> While these two metrics are essentially independent of each other, there are some practical dependencies between them. First of all, an order similarity score greater than zero naturally demands that there be some shared recipes between the two versions, i.e. the extent similarity score also has to be greater than zero. Second, a decrease in extent similarity tends to also decrease the level of order similarity, since an increase in the number of non-matching recipes increases the likelihood of these non-matching recipes interrupting sequences of matching ones.

<sup>3</sup> Since the *Jaro distance* and the *Jaro–Winkler distance* (Jaro 1989; Winkler 1990), commonly used for string comparison, have been formulated specifically for the comparison of natural-language words and assume a string length typical to natural languages, they are not appropriate for the present purpose.

<sup>4</sup> Since all of the practical implementations of these metrics—of which this analysis uses the one provided by the `stringdist` package (<<http://cran.r-project.org/web/packages/stringdist/>>) written by Mark van der Loo for the *R* statistics software (<<http://www.r-project.org/>>)—are based on the

	Harley 4016	Harley 279	Additional 5467	Ashmole 1439	Douce 55	Cosin V.iii.11 A
Harley 4016	—	Overall similarity: 9% Shared: 79/182 (43%) Consec.: 16/79 (20%) Lv: 251 (1.38) Lv <sub>s</sub> : 78 (0.99)	Overall similarity: 40% <b>Shared: 165/182 (91%)</b> Consec.: 73/165 (44%) Lv: 180 (0.99) Lv <sub>s</sub> : 164 (0.99)	Overall similarity: 9% Shared: 87/182 (48%) Consec.: 16/87 (18%) Lv: 260 (1.43) Lv <sub>s</sub> : 88 (1.01)	Overall similarity: 38% <b>Shared: 168/182 (92%)</b> Consec.: 69/168 (41%) Lv: 182 (1) Lv <sub>s</sub> : 167 (0.99)	Overall similarity: 79% <b>Shared: 156/182 (86%)</b> <b>Consec.: 144/156 (92%)</b> Lv: 51 -2 (0.28) Lv <sub>s</sub> : 21 -3 (0.13)
Harley 279	Overall similarity: 6% Shared: 79/258 (31%) Consec.: 16/79 (20%) Lv: 251 (0.97) Lv <sub>s</sub> : 78 (0.99)	—	Overall similarity: 12% Shared: 92/258 (36%) Consec.: 31/92 (34%) Lv: 248 (0.96) Lv <sub>s</sub> : 85 (0.92)	<b>Overall similarity: 96%</b> <b>Shared: 248/258 (96%)</b> <b>Consec.: 248/248 (100%)</b> Lv: 33 -1 (0.13) Lv <sub>s</sub> : 6 -1 (0.02)	Overall similarity: 12% Shared: 92/258 (36%) Consec.: 31/92 (34%) Lv: 248 (0.96) Lv <sub>s</sub> : 85 (0.92)	Overall similarity: 5% Shared: 80/258 (31%) Consec.: 13/80 (16%) Lv: 250 (0.97) Lv <sub>s</sub> : 77 (0.96)
Additional 5467	Overall similarity: 40% <b>Shared: 165/181 (91%)</b> Consec.: 73/165 (44%) Lv: 180 (0.99) Lv <sub>s</sub> : 164 (0.99)	Overall similarity: 17% Shared: 92/181 (51%) Consec.: 31/92 (34%) Lv: 248 (1.37) Lv <sub>s</sub> : 85 (0.92)	—	Overall similarity: 17% Shared: 99/181 (55%) Consec.: 31/99 (31%) Lv: 259 (1.43) Lv <sub>s</sub> : 93 (0.94)	<b>Overall similarity: 98%</b> <b>Shared: 180/181 (99%)</b> <b>Consec.: 177/180 (98%)</b> Lv: 7 (0.04) Lv <sub>s</sub> : 2 (0.01)	Overall similarity: 39% <b>Shared: 164/181 (91%)</b> Consec.: 70/164 (43%) Lv: 173 (0.96) Lv <sub>s</sub> : 159 (0.97)
Ashmole 1439	Overall similarity: 6% Shared: 87/269 (32%) Consec.: 16/87 (18%) Lv: 260 (0.97) Lv <sub>s</sub> : 88 (1.01)	<b>Overall similarity: 92%</b> <b>Shared: 248/269 (92%)</b> <b>Consec.: 248/248 (100%)</b> Lv: 33 -1 (0.12) Lv <sub>s</sub> : 6 -1 (0.02)	Overall similarity: 12% Shared: 99/269 (37%) Consec.: 31/99 (31%) Lv: 259 (0.96) Lv <sub>s</sub> : 93 (0.94)	—	Overall similarity: 12% Shared: 99/269 (37%) Consec.: 31/99 (31%) Lv: 259 (0.96) Lv <sub>s</sub> : 93 (0.94)	Overall similarity: 5% Shared: 88/269 (33%) Consec.: 13/88 (15%) Lv: 260 (0.97) Lv <sub>s</sub> : 86 (0.98)
Douce 55	Overall similarity: 37% <b>Shared: 168/184 (91%)</b> Consec.: 69/168 (41%) Lv: 182 (0.99) Lv <sub>s</sub> : 167 (0.99)	Overall similarity: 17% Shared: 92/184 (50%) Consec.: 31/92 (34%) Lv: 248 (1.35) Lv <sub>s</sub> : 85 (0.92)	<b>Overall similarity: 96%</b> <b>Shared: 180/184 (98%)</b> <b>Consec.: 177/180 (98%)</b> Lv: 7 (0.04) Lv <sub>s</sub> : 2 (0.01)	Overall similarity: 17% Shared: 99/184 (54%) Consec.: 31/99 (31%) Lv: 259 (1.41) Lv <sub>s</sub> : 93 (0.94)	—	Overall similarity: 37% <b>Shared: 167/184 (91%)</b> Consec.: 68/167 (41%) Lv: 175 (0.95) Lv <sub>s</sub> : 163 (0.98)
Cosin V.iii.11 A	<b>Overall similarity: 85%</b> <b>Shared: 156/169 (92%)</b> <b>Consec.: 144/156 (92%)</b> Lv: 51 -2 (0.3) Lv <sub>s</sub> : 21 -3 (0.13)	Overall similarity: 8% Shared: 80/169 (47%) Consec.: 13/80 (16%) Lv: 250 (1.48) Lv <sub>s</sub> : 77 (0.96)	Overall similarity: 41% <b>Shared: 164/169 (97%)</b> Consec.: 70/164 (43%) Lv: 173 (1.02) Lv <sub>s</sub> : 159 (0.97)	Overall similarity: 8% Shared: 88/169 (52%) Consec.: 13/88 (15%) Lv: 260 (1.54) Lv <sub>s</sub> : 86 (0.98)	Overall similarity: 40% <b>Shared: 167/169 (99%)</b> Consec.: 68/167 (41%) Lv: 175 (1.04) Lv <sub>s</sub> : 163 (0.98)	—

Table 13.1: The overall similarity of the version pairs, with the number (and %) of shared recipes and the number (and %) of recipes occurring consecutively in both, along with the raw and normalised Levenshtein (Lv) distance, calculated separately over the entire collection and just over the shared recipes (Lv<sub>s</sub>).

sequence than is appropriate in this context, in addition to which they are also restricted to comparing two versions. Because of these limitations, the similarity of recipe order was also quantified by simply counting the proportion of shared recipes which occur in the same local sequential context, i.e. for which either the preceding or following (or both) recipe is the same in all versions. This proportion was calculated simply by dividing the number of recipes occurring in the same local context in all versions by the total number of recipes shared by the versions being compared.

The results of these quantitative analyses are shown in *Table 13.1*, with both absolute counts and percentages given for the number of shared recipes and the number of shared recipes occurring in the same sequential context.<sup>5</sup> The overall similarity score shown here is simply a product of the similarities of extent (percentage of shared recipes) and of order (percentage of shared recipes occurring in the same sequential context). The Levenshtein distance between the two versions has been calculated for both the entire collection as a whole ( $L_v$ ) and just for the recipes shared between the two versions ( $L_{v_s}$ ).<sup>6</sup> Since the Levenshtein and the Damerau–Levenshtein distance are the same in most cases, they have not been listed separately but rather a difference between the Levenshtein and D–L distances has been indicated by a modifier (e.g. -2) following the raw Levenshtein distance.<sup>7</sup> In order to highlight pairs exhibiting a high degree of similarity, all values for the number of shared recipes and for recipes occurring in the same local context exceeding 75% have been printed in bold face.

Examining the values presented in *Table 13.1* quickly reveals that there are in fact three pairs of versions that exhibit a significant degree of mutual similarity: MSS Additional 5467 (MS Ad) and Douce 55 (MS D), MSS Harley 279 (MS H279) and Ashmole 1439 (MS As), and MSS Harley 4016 (MS H4016) and Cosin V.iii.11 A (MS C). All of these pairs not only share the vast majority of their recipes, but also present them in a very similar order. Thus the *Potage Dyvers* family would seem to consist of three major versions in terms of textual content and structure, each surviving in two manuscripts with minor variations. Of these three pairs, the closest resemblance is found between MSS Ad and D, which are practically identical, the latter containing all but one of the recipes in the former in practically identical order, while adding further four recipes that do not occur in the former.

---

comparison of character strings (instead of sequences like a list of recipe identifiers), the list of recipes in each version was converted into a sequence of Unicode characters by replacing each recipe identifier (of the format *PD n*, where *n* is a number between 1 and 371) with a Unicode character with a decimal character code of  $n + 200$  (e.g. &#241;). These strings, in which each unique Unicode character thus represents one of the 371 unique recipes found in the *Potage Dyvers* collection were then used as the string vectors for the `stringdist` algorithm.

<sup>5</sup> The reference version being compared to the other versions is indicated in the leftmost column, and the different versions it is compared to in the top row. While the absolute values for the metrics are symmetrical between the two versions being compared, the percentages and normalised values given in parentheses are calculated with reference to the version in the left column, and are thus different for the inverse versions of each pair.

<sup>6</sup> The Levenshtein distance calculated for only the shared recipes ( $L_{v_s}$ ) ignores all of the non-shared recipes interposed between the shared ones, and is therefore likely to be proportionally much lower than the one for all recipes.

<sup>7</sup> This difference between the two metrics indicates the number of recipes that occur consecutively in both collections but in inverted order, as this situation involves two operations in the Levenshtein metric (deletion + addition) but only a single one in the Damerau–Levenshtein metric (transposition).

#							Shared / Consec.
3			Ad	D		H4016	164 / 67
3			Ad	D	C		163 / 66
3				D	C	H4016	155 / 60
3			Ad		C	H4016	152 / 64
4			Ad	D	C	H4016	151 / 58
3	As		Ad	D			99 / 31
3		H279	Ad	D			92 / 31
3	As	H279	Ad				91 / 31
4	As	H279	Ad	D			91 / 31
3	As	H279		D			91 / 31
4	As		Ad	D	C		88 / 11
3	As		Ad		C		88 / 11
3	As			D	C		88 / 11
3	As		Ad			H4016	87 / 12
4	As		Ad	D		H4016	87 / 12
3	As			D		H4016	87 / 12
3	As	H279			C		80 / 13
5	As	H279	Ad	D	C		80 / 11
4	As	H279	Ad		C		80 / 11
4		H279	Ad	D	C		80 / 11
3		H279	Ad		C		80 / 11
4	As	H279		D	C		80 / 11
3		H279		D	C		80 / 11
3	As	H279				H4016	79 / 16
3		H279	Ad			H4016	79 / 12
4	As	H279	Ad			H4016	79 / 12
5	As	H279	Ad	D		H4016	79 / 12
4		H279	Ad	D		H4016	79 / 12
4	As	H279		D		H4016	79 / 12
3		H279		D		H4016	79 / 12
3	As				C	H4016	78 / 12
5	As		Ad	D	C	H4016	78 / 10
4	As		Ad		C	H4016	78 / 10
4	As			D	C	H4016	78 / 10
4	As	H279			C	H4016	70 / 12
3		H279			C	H4016	70 / 12
5	As	H279	Ad		C	H4016	70 / 10
5		H279	Ad	D	C	H4016	70 / 10
4		H279	Ad		C	H4016	70 / 10
5	As	H279		D	C	H4016	70 / 10
4		H279		D	C	H4016	70 / 10
6	As	H279	Ad	D	C	H4016	70 / 10

**Table 13.2:** *The number of shared recipes and the number of recipes occurring in the same sequence in each of the versions for groups of three to six versions, in decreasing order of shared and co-sequential recipes.*

The longest versions of the family, MSS H279 and As, which contain about 50 per cent more recipes than the other versions both share more than 90 per cent of these recipes with each other, again in a virtually identical order, both adding a handful of unique recipes to the shared stock of 245 recipes. The least similar of the three discernible pairs is made up by MSS H4016 and C, which share only 86 and 92 per cent, respectively, of their recipes with each other, and also display more fluctuation in the order of these shared recipes.

Furthermore, if we look at the pairs that are very similar in extent but not



in order, we can see that these are all permutations of MSS H4016, Ad, D, and C, showing that these two pairs share the majority of recipes between them, but organise them very differently.<sup>8</sup> The H279–As pair, on the other hand, differs from the rest in both respects, containing a large number of recipes not found in the other versions and omitting roughly half of the recipes shared by the other four versions. Looking at the larger sets of versions in *Table 13.2*, we can see that MSS H4016, Ad, D and C share more recipes among them than any set of three or more versions that is not its subset, and share 50% more recipes among themselves than any set containing either MS H279 or As. At the other end of the table we can see that the minimum degree of similarity is found between the H279–As and H4016–C pairs which do not share any recipes beyond the common core shared by all versions, while the As–H279 and Ad–D pairs do share three recipes (one of which does not occur in MS As) which do not occur in either MS C or MS H4016, possibly indicating ‘cross-contamination’ or a shared external source somewhere along their textual history.

Looking at the two Levenshtein distances (for all recipes and only shared recipes) for the three types of recipe pairs—similar in both extent and order, similar only in extent, and similar in neither—we can see that the Levenshtein distance is a good predictor of order similarity, as would be expected, but does not react to the same recipes occurring in a different order since it does not differentiate transpositions from additions and deletions. Unfortunately, even the standard Damerau-Levenshtein distance, which does account for transpositions separately as a single operation, is of no use in the case of recipe sequences because it only accounts for transpositions between *adjacent* recipes. A more suitable metric for accounting for the overall similarity, taking into account both extent and order, would be an unrestricted variant of the Damerau-Levenshtein distance, but unfortunately the practical implementation of a Damerau-Levenshtein variant with transposition over unrestricted distances is extremely complicated and no practical implementations seem to be generally available.

Taking a more structured look at the recipes shared between all of the pairwise permutations of the six versions, *Figure 13.1* illustrates the spatial distribution of the recipes shared between each pair. The same three pairs distinguished above on quantitative grounds are immediately discernible here—especially the very similarly ordered Ad–D and H279–As pairs, which occur as almost entirely black. The four other pairs which share extent but not order are distinguishable as mainly dark grey with short groups of recipes occurring across both versions. The H4016–C pair can be seen to coincide more closely in its latter part, the beginning part being more particular to each version. Looking at the less similar pairs, we can see that the shared recipes are spread quite evenly across the length of the collection, pointing towards the common core of 66 recipes shared by all versions consisting of a variety of common recipes thoroughly integrated with recipes particular to specific pairs of versions or even to an individual version. However, there do seem to be some patterns in their occurrence. While the H4016–C pair contains the most variation in the beginning of the collection, it seems that in general the initial parts of the collections would seem to contain more of the universal recipes

<sup>8</sup> For example C is in fact more similar to MSS Ad and D in terms of its extent, but the completely different organisation of those recipes means that the overall degree of similarity with these versions is considerably lower than with MS H4016.

than their final parts, although MSS Ad and C also seem to contain a concentration of recipes shared with the other versions around their middle section.

Based on this macro-level examination of the relative extent and ordering of the recipes in each of the six surviving members of the *Potage Dyvers* family of recipe collections, they would thus seem to constitute three pairs, of which two are more closely related to each other, and the third less so. Since there is no single ‘natural’ principle for selecting and ordering the recipes, each reordering of the recipes can be seen as essentially idiosyncratic, making it very unlikely that two copyists would spontaneously decide upon an identical selection and order of recipes unless influenced by a shared ancestor. This means that we can use these similarities and differences in the extent and order of recipes to construct a tentative transmissional history for the different versions, postulating shared ancestor versions for the closely related pairs of recipes, and a further one for the ancestors of the two related pairs. These ancestors and the resulting family tree—with  $\alpha$  representing the highly hypothetical ‘original’ version, which may well never have existed—is presented graphically in *Figure 13.2*. While the organisation of recipes within the collection is naturally only one of the possible bases for establishing textual filiation, this family tree—even if it does not necessarily represent the actual copying sequence of the versions—does provide a useful shorthand for referring to the different groups of versions and a starting point for the more detailed analysis undertaken below in section 13.2.

## 13.2 Centrality and distribution of the recipes

In order to shed more light on the structural relationships of the six manuscript versions of the *Potage Dyvers* family and to provide clues about the textual history of the three structurally different pairs of manuscript versions, this section first takes a look at the differing *centrality* (see subsection 2.1.4) of recipes and recipe groups, i.e. the number of versions in which they occur, and then examines in more detail the differences in the ordering of recipes within the six versions of *Potage Dyvers* and the degree to which clusters of recipes occur in the same order between the different groups of versions. As a purely quantitative starting point, *Table 13.3* divides the 371 unique recipes into categories based on the number of manuscript versions in which they occur.<sup>9</sup> For a more detailed representation distinguishing between specific combinations of versions, *Figure 13.3* uses an angular or ‘embattled’ unweighted Edwards-Venn diagram (Edwards 1989) of six sets, with each set representing a single manuscript version and the areas of overlap representing recipes shared between the different versions.<sup>10</sup> The solid

<sup>9</sup> The extra recipe shared by all six versions in MSS H4016, H279 and As is explained by the fact that both versions two occurrences of a unique recipe (PD 49 in MS H4016 and PD 71 in MSS H279 and As).

<sup>10</sup> Since this type of diagram includes all logically possible combinations of the different versions, even if there are no recipes contained only in that combination of versions and the areas are not weighted for the number of recipes covered by them, it is not the ideal representation for this data. Unfortunately there do not seem to be any available visualisation algorithms that could produce a weighted Euler diagram with six sets, and the available visualisations even for unweighted Venn diagrams of six sets are very limited. *Figure 13.3* has been produced using the *Vennerable* visualisation package (<<http://r-forge.r-project.org/projects/vennerable/>>) for the *R* statistical computing

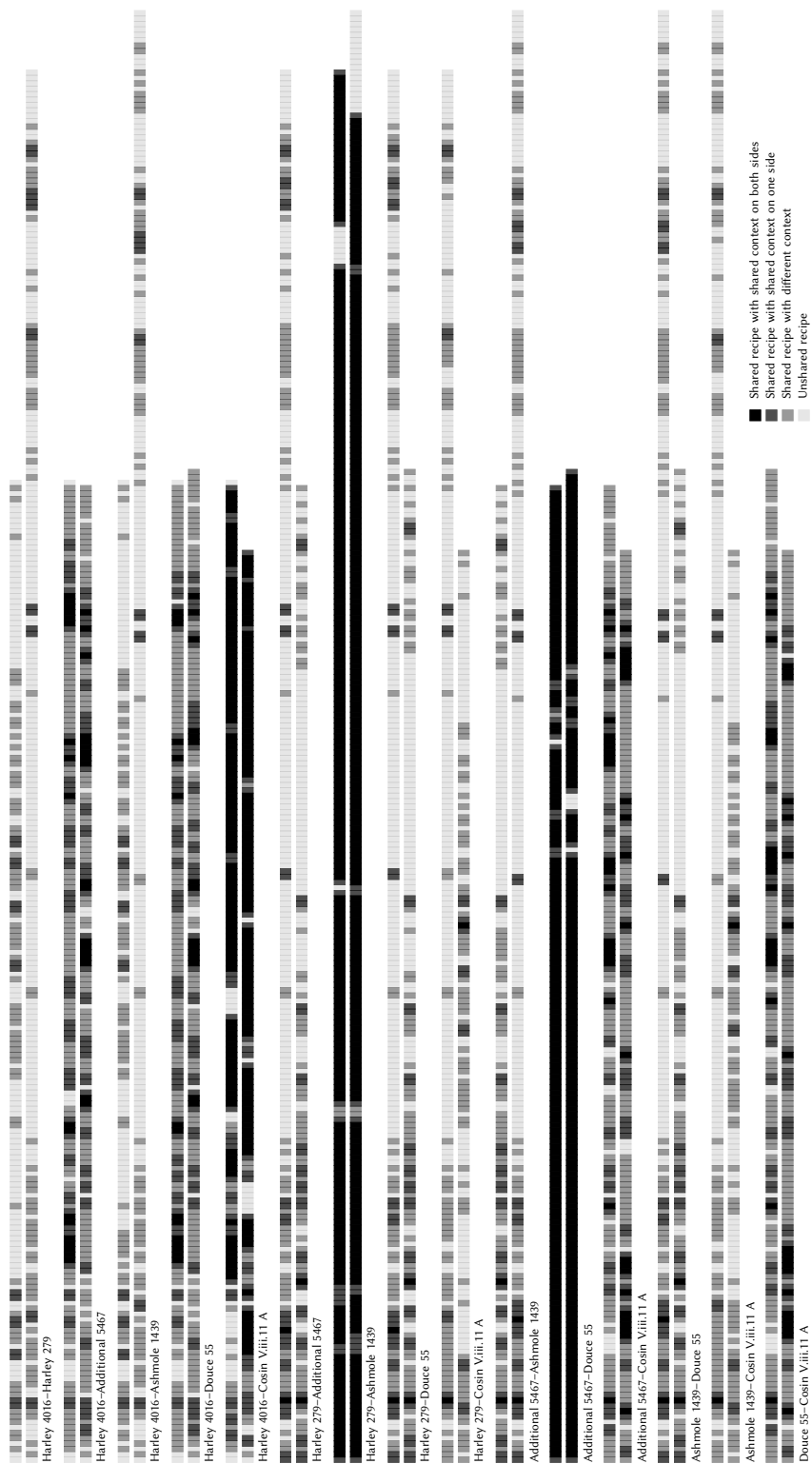
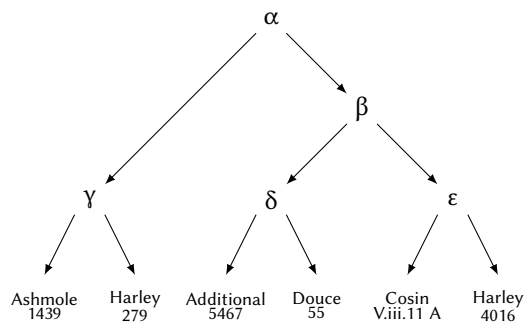


Figure 13.1: Structural similarity between the different pairs of Potage Dyvers versions in terms of their selection and ordering of recipes.



**Figure 13.2:** *The transmissional relationships suggested by a quantitative analysis of the selection and ordering of recipes in the six surviving versions of the Potage Dyvers family of recipe collections.*

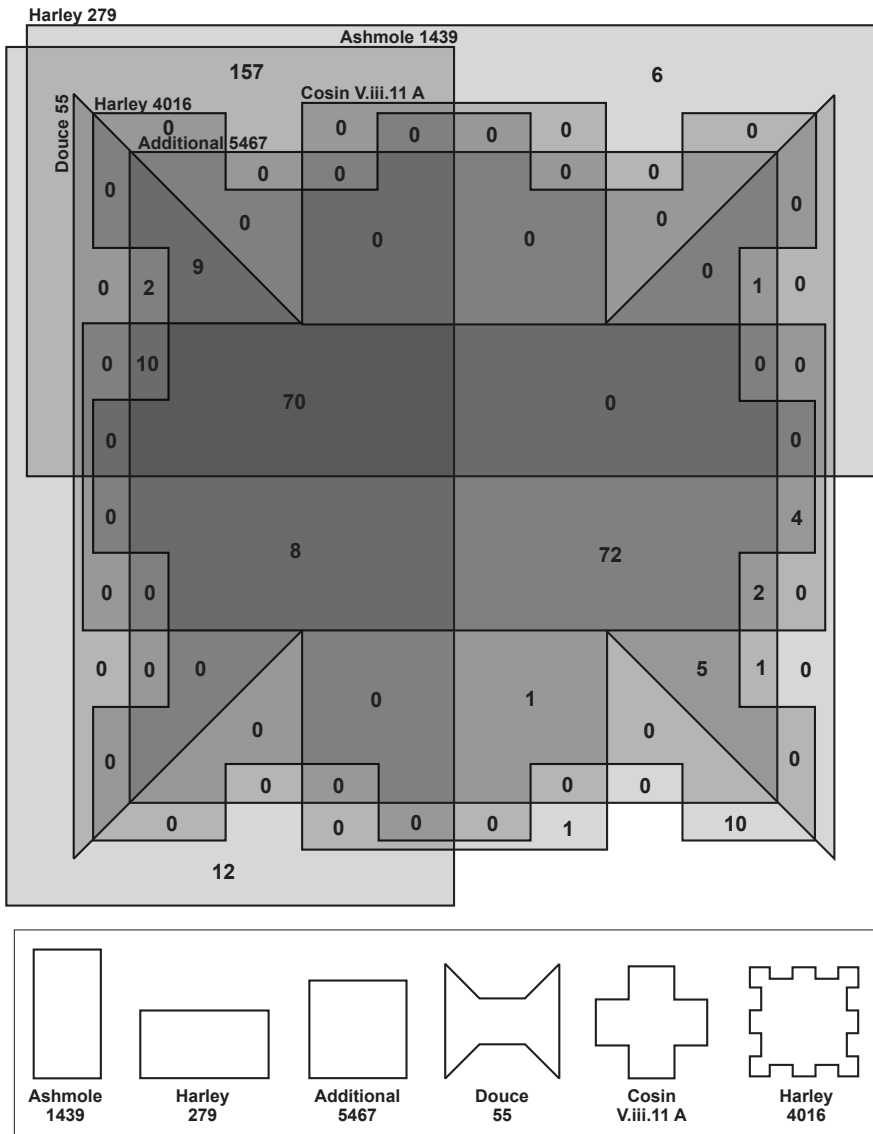
No. of versions	Harley 4016	Harley 279	Additional 5467	Ashmole 1439	Douce 55	Cosin V.iii.11
6	71 (39%)	71 (28%)	70 (39%)	71 (26%)	70 (38%)	70 (41%)
5	17 (9%)	19 (7%)	27 (15%)	27 (10%)	27 (15%)	18 (11%)
4	75 (41%)	2 (1%)	75 (41%)	2 (1%)	75 (41%)	73 (43%)
3	9 (5%)	1 (0%)	8 (4%)	0 (0%)	11 (6%)	7 (4%)
2	0 (0%)	158 (61%)	1 (1%)	158 (59%)	1 (1%)	0 (0%)
1	10 (5%)	7 (3%)	0 (0%)	11 (4%)	0 (0%)	1 (1%)
Total	182	258	181	269	184	169

**Table 13.3:** *The recipes in each version, categorised according to their centrality to the Potage Dyvers family seen as a discourse colony, as determined by the number of manuscript versions they occur in.*

core of the *Potage Dyvers* family of collections is made up of the 70 recipes that occur in all of the six versions, along with the 27 occurring in all but one version, having been either lost due to damage or accidentally omitted from MS H279, C or H4016. While most likely originating from a shared source, these ‘core recipes’ do not constitute a separate structural unit in any of the versions, but have been thoroughly integrated with recipes from other sources, being more or less evenly interspersed with recipes occurring in only some of the *PD* versions (see *Figure 13.6* on page 621).

The degree of variance in the organisation of the six versions is reflected by the fact that despite the relatively large number of recipes shared by all versions, no sequence of consecutive recipes longer than two occurs in all of the six versions, the five pairs of recipes that occur together in all versions being PD 11–2, PD 48–9, PD 51–2, PD 54–5 and PD 71–2.<sup>11</sup> In terms of their content, all of these pairs are for two variations of the same basic dish, which explains why they have been kept

software (<<http://www.r-project.org/>>) and manually redrawn for clearer grayscale representation.  
<sup>11</sup> The first recipe of this last pair not only occurs in all of the versions, but actually occurs twice in MSS H279 and As with minor variations in its content.



PD versions	Recipe sequences
As / H279 / Ad / D / C	PD 11–PD 12–PD 13
Ad D C H4016	PD 92–PD 93–PD 94–PD 95–PD 96–PD 97–PD 98 PD 133–PD 134–PD 135–PD 136–PD 137 PD 106–PD 107–PD 108 PD 152–PD 153–PD 154 PD 157–PD 158–PD 159 PD 160–PD 161–PD 162
As / H279 / Ad / D	PD 33–PD 34–PD 35
Ad / D / C	PD 109–PD 110–PD 111–PD 112–PD 113 PD 47–PD 48–PD 49 PD 85–PD 86–PD 87
H4016 / Ad / C	PD 133–PD 134–PD 135–PD 136–PD 137–PD 184–PD 138–PD 139 PD 140–PD 141–PD 147
H4016 / Ad / D	PD 66–PD 67–PD 68–PD 69 PD 106–PD 107–PD 108–PD 109

**Table 13.4:** *The sequences of more than two consecutive recipes that occur in the same order in more than two different versions of the Potage Dyvers family.*

together in all of the manuscript versions despite the widely differing extent and organisation of the major  $\beta$  and  $\gamma$  versions and the extensive differences in the organisation of the  $\delta$  and  $\epsilon$  versions.<sup>12</sup> The clear division of the six versions into three pairs, seen in the above summary analysis, is also evident in the number and length of recipe groups shared by sets of versions. There are no significant groups of consecutive recipes shared by five of the versions, apart from the ones shared by all six,<sup>13</sup> and no set of three versions shares a significant number of substantial groups in addition to those shared by four of the versions.<sup>14</sup>

As can be seen in *Table 13.4*, the largest set of versions which shares recipe groups of any significant length is the same one identified as a set of closely related pairs in *Table 13.2* and as an area of *Figure 13.3* containing a large number (72) of recipes, namely the  $\beta$  group. The longest continuous series of recipes that occurs in all of these four collections includes recipes PD 82–88 and occurs near the beginning of MSS H4016 and C and in the middle of MSS Ad and D, providing instruction for the preparation of seven different kinds of cold sauces.<sup>15</sup> Two further sequences which occur in either exactly or nearly the same order in these

<sup>12</sup> The first pair consists of a recipe for simple almond milk and a recipe for a kind of ‘almond cheese’ made by boiling almond milk, curdling it with vinegar, and draining it to produce what is essentially almond cottage cheese, while the second pair consists of two versions—a simpler and a more complex one—for a meat paste pressed into a form and sliced for serving. The third pair consists of two wine-based *soppes*—broths poured over pieces of toast—one thickened with ground almonds, and the fourth one of two versions of *Leach Lombard*, slices of sweet paste based on dates in the first version and on honey stiffened with cooked egg yolks in the second.

<sup>13</sup> All of the versions except for MS H4016 share a single additional recipe after the first universal pair, namely PD 13, while all of the versions except for MS C have the pair PD 22–3 occurring together.

<sup>14</sup> As is also indicated by *Table 13.3*, all of the sequences of consecutive recipes shared by three versions—except for a single pair (PD 16 & 40, shared between MSS H4016, H279 and As) are shared between some (partial) permutation of the  $\beta$  group of versions, which were found to be closely related in the above analysis.

<sup>15</sup> These recipes do not occur at all in MS H279 but all but one of them (PD 93, “Gauncele Sauce”) do occur as a part of a longer series of recipes in MS As, although in a different order.

four versions are constituted by recipes PD 133–7, which occur as a continuous sequence near the end in all of the four versions, and recipes PD 152–4 & 157–62, which occur as three clusters of three consecutive recipes in close proximity in all of the versions, interspersed by recipes PD 155–156 in MSS Ad and D, and by recipes PD 164, 165, 173 and 181 in MSS H4016 and C.<sup>16</sup> However, the proximity of these groups of three in all four collections seems coincidental: while the first two of them are internally thematically consistent (PD 152–5 being for pastries fried in oil, PD 157–9 for hot soups, and 160–2 for two fish stews and one for sweetmeat made of figs), they do not form a thematic group as a whole.<sup>17</sup>

In addition to these sequences, these four versions also share a number of recipe pairs or triplets, many of which are a part of a larger sequence with a similar but not identical organisation in all four versions. For example the sequence of recipes PD 106–8 is actually a part of a longer sequence of 16 recipes for roasted birds and rabbits, occurring in slightly different order in all of the four versions.<sup>18</sup> In addition to these slightly longer sequences, there are also 13 recipe pairs that occur as items in all of these four versions (PD 6–7, 9–10, 36–7, 61–2, 71–2, 78–9, 86–7, 110–1, 118–9, 125–6, 138–9, 140–1, and 164–5), many of them being parts of larger wholes which occur in all of these four versions with minor variations in order, and in different place in the collection.<sup>19</sup> In addition to the sequences shared by these four closely related versions, there is a single consecutive pair of recipes shared by MSS H4016, H279, As, and C (but *not* by MSS Ad and D, where the two recipes occur separated by five other recipes), consisting of recipes for omelets flavoured (and coloured) by tansy (PD 174) and by beef or veal (PD 180).

Of the longer sequences shared by three of the four closely related versions, most are extensions of the shorter sequences shared by all four versions, mentioned in a note above. For example the longest sequence (5 consecutive recipes) shared by MSS Ad, D and C consists of recipes PD 109–13, which is a subset of the loosely shared sequence of bird recipes described above, and the longest sequence (8 consecutive recipes) shared by MSS H4016, Ad and C, consisting of recipes PD 133–7, 184 & 138–9, is a subset of the sequence of fish described above. Others are sequences which are shared by all of the four versions, but with slightly different order in one of the versions, as in the case of PD 66–9 and 106–9, which occur in the same order in MSS H4016, Ad and D, but with a single inversion of two recipes

<sup>16</sup> The order of the three clusters is also reversed between the  $\delta$  and  $\epsilon$  groups.

<sup>17</sup> The two recipes interposed between PD 154 and 157 in MSS Ad and D are for fried and roasted sweet dishes, continuing the theme of the preceding group of three recipes, and the first of the four recipes interposed between recipes PD 162 and PD 157 in MSS H4016 and C is for a paste of dried fruit and the rest for two thickened broths of fish and a jelly, which connects it to the surrounding groups, which are also for fish soups and sweetmeats.

<sup>18</sup> Unlike in many other cases, even the two pairs of recipes identified as very close matches do not internally match each other; in MSS Ad and D the sequence is identical except for PD 114 missing from MS Ad, while MSS H4016 and C invert the order of PD 112 and 113, and furthermore insert here a recipe for stewed partridge occurring elsewhere in the other four versions (PD 28, added between PD 108 and 109 in MS C and between PD 109 and 110 in MS H4016).

<sup>19</sup> For example PD 138–141 belong to a group of recipes for various fish, which occurs in roughly the same position near the end of the collection in all four versions, but with minor variations in the order of the recipes and with some versions including recipes for fish that in other versions occur elsewhere in the collection. In MSS Ad and D this sequence contains 22 or 21 recipes, while MSS H4016 and C extend it to 40 and 38 recipes by including other fish and seafood recipes occurring elsewhere in the other versions.

and one recipe lost to damage in MS C.

### 13.2.1 Comparing recipe order in pairs of versions

It is only at the level of the ‘sibling’ pairs of versions— $\gamma$ ,  $\delta$  and  $\epsilon$ —identified above that we see significant stretches of recipes shared in the exact same order, and even these pairs contain numerous minor differences. While the greatest absolute number of recipes occurring in the same local context in two versions (247) and the longest sequence of consecutive recipes occurring in the same sequence in two versions (115 consecutive recipes, numbered 108–222 in MS H279 and 107–221 in MS As) is found in the  $\gamma$  pair, the most uniform pair in relative terms is the  $\delta$  pair, as 98 or 96 per cent of the 181 or 184 recipes in MS Ad and MS D occur in the same local context in both versions, the longest sequence of consecutive recipes occurring in identical order being the first 113 recipes of both collections which account for almost two thirds of each collection. The least similar pair in terms of both the longest identically ordered sequence (43) and the total number of recipes occurring in the same local context (144) is the  $\epsilon$  group.<sup>20</sup> MSS H4016 and C are also the only versions that contain sequences—located mainly in the initial part of these versions—shared between the whole  $\beta$  group that are not subsets of longer sequences shared between the two ‘sibling’ versions, the initial parts of these MSS thus exhibiting as much variation between themselves as between them and the  $\delta$  group. The final part of this section will examine the specific differences within the three pairs of collections one at a time, from the most to the least similar.

#### MSS Additional 5467 and Douce 55 ( $\delta$ group)

Even a cursory examination of the  $\delta$  group (MSS Ad and D) in *Figure 13.1* quickly reveals that the differences between them are restricted to roughly the third quarter of the collection, and consist of the omission of PD 114—a recipe for roasted *brew* (possibly ‘whimbrel’, a species of curlew)—and recipes PD 122–124—recipes for roasted kid, veal and venison (inserted between the recipes for game birds and two recipes for chicken)—from MS Ad, and the most likely accidental omission of PD 184—a recipe for boiled bream or roach—from MS D.<sup>21</sup> Apart from these differences, which also explain the slightly differing lengths of the two collections, the two versions of the  $\delta$  group are identical in terms of both their extent and organisation.

#### MSS Harley 279 and Ashmole 1439 ( $\gamma$ group)

The second of these groups in terms of internal cohesion is the  $\gamma$  group (MSS H279 and As), which is almost 50 per cent longer than the other groups and clearly the

<sup>20</sup> Both the total number and the longest sequence include two recipes (PD 141 and 149) that have been divided into two separate recipes in MS H4016 because they provide two alternative versions of the recipe.

<sup>21</sup> The differences are described in terms of omissions rather than additions for the simple reason that all of them occur in a very similar context in the  $\epsilon$  group and are thus likely to have been a part of a common ancestor. While the omission of the recipes in MS Ad could be either intentional or accidental, the fact that the title of the recipe omitted from D, *Breme or roche buillez*, is erroneously applied to the following recipe (which is in fact for *roasted* bream, as indicated by the titles of the other versions) makes it likely that the recipe was omitted by accident.



most distinctive of the three groups in terms of both its extent and organisation. Unlike in the  $\delta$  group, the differences are not limited to a specific part of the collection, although the greatest variation—especially in terms of extent—is located near the end of the collection, as can be seen in *Figure 13.1*. The differences in the initial part of the collection involve the transposition of recipes either ‘*in situ*’ as in the case of recipes PD 202 and 203 (recipes number 65 and 66 in the collection) which occur in inverted order in the two versions, or through transferral from one place to another in the collection, as in the case of PD 47 and 56 (recipes 22–23 in MS As and 31–32 in MS H279).<sup>22</sup>

The first difference in the extent of the two versions is found near the middle of the collection, where a unique recipe for *Eel in Sorre* (recipe 107, PD 365) has been added to MS H279 in a rather illogical place between recipes for a *Meat Rapee* (PD 241) and for a primrose pudding (PD 242).<sup>23</sup> The first major difference near the end of the collection is actually the result of damage to MS As, involving the loss of two folia between ff. 40 and 41, which most likely held the seven recipes occurring at this point in MS H279, as well as the missing beginning of MS As recipe 222 (PD 329). These eight recipes (including the almost entirely missing PD 329) take up two full folia in MS H279—which is the number of leaves missing from MS As at this point, making it likely that these were the only recipes originally contained at this point.<sup>24</sup> This means that the two versions were originally considerably more similar than the figures in *Table 13.1* suggest. The only major difference in the extent or organisation of these two versions that is not due to later accidents is the inclusion of 19 recipes for sauces at the end of MS As (recipes 251–69). None of these recipes occur in MS H279, but seven of them are found in all of the  $\beta$  versions, although in varying order and interspersed with additional sauce recipes.

### MSS Cosin V.iii.11 A and Harley 4016 ( $\epsilon$ group)

The least similar of the three closely related pairs—in terms of both extent and organisation—is  $\epsilon$ , whose versions differ significantly from each other especially in the first third of the collection. While some of this variation is caused by the two folia that have been lost from between ff. 3 and 4 and ff. 9 and 10 of MS C, much of it seems to have been present already when the manuscripts were copied. In terms of extent, the major differences—apart from the recipes lost to damage—include the omission of six recipes from between recipes 59 and 60 in MS H4016, suggesting

<sup>22</sup> These two recipes are for two different versions of *Brawn in Pevorade* which occur before a sequence of soups in MS As but near its end (preceding a recipe for *Oil Sops*) in MS H279. Since these two recipes have been transposed as a pair, they are shown in *Figure 13.1* as sharing one half of their context (i.e. each other) in both of the versions. It is impossible to tell whether the recipes have been accidentally transposed by the scribe of MS H279 (possibly as a result of accidentally skipping them first) or moved into a more reasonable position by the scribe of MS As.

<sup>23</sup> Since these two manuscripts share a large number of recipes that do not occur in any other version, it is equally possible that this recipe was included in their common ancestor and omitted from MS As.

<sup>24</sup> For some reason, Hieatt (2004: 27) counts the number of recipes missing at this point as 11, giving the number of recipes in this section of MS H279 as 44 while it is—according to the calculation of both myself and Austin (1888)—actually 41. She also seems to miss the irregularity in the collation of the MS at this point (see subsection 9.2.4), as she argues the most likely reason for the omission to lie in a faulty exemplar used by the Ashmole scribe.

	MS Cosin V.iii.11 A	MS Harley 4016
Added to	343 <sup>29</sup>	PD 344, PD 345, PD 346, PD 347, PD 348, PD 349, PD 350, PD 351, PD 352, PD 353
Omitted from	(PD 8), PD 22, (PD 23), PD 40, (PD 45), PD 58, (PD 69), (PD 89), (PD 21), (PD 24), (PD 26), (PD 76), (PD 77)	PD 1, PD 3, PD 27, PD 29, PD 33, PD 35, PD 85, PD 99, PD 102, PD 103, PD 129

**Table 13.5:** *The differences between MSS C and H4016 indicated as lists of recipes omitted from or added to either version.*

that it might have been copied from a damaged exemplar.<sup>25</sup> Although the differing organisation of the two versions makes it difficult to determine with certainty the recipes that have been lost with the missing folia, evidence provided by the undamaged table of contents in MS C would seem to indicate that the missing folio between ff. 3 and 4 contained recipes PD 40, 8, 23, 45, 25, and 70, along with an unidentified recipe for “Chewettes”,<sup>26</sup> and the one between ff. 9 and 10 contained recipes PD 21, 24, 26, 69, 89, 76, and 77.<sup>27</sup> However, the fact that the order of the table of contents does not seem to correspond exactly with the order of the recipes in the collection itself makes this somewhat uncertain. Apart from these larger groups of recipes, the differences in extent between the two versions consist mostly of individual recipes or pairs of recipes that have been added to or omitted from one of the two versions. Judging from the occurrence of these differentiating recipes in the rest of the versions, MS H4016 would seem to be the more ‘progressive’ of the two, since it contains a number of unique recipes not attested elsewhere, whereas the recipes attested in MS C but not in H4016 are also found in most of the other versions and are thus more likely omissions of the Harley scribe rather than innovations of the Cosin one. *Table 13.5* contains a representation of the differences in the extent of the two  $\epsilon$  versions.<sup>28</sup>

In terms of differences in their organisation, the  $\epsilon$  versions are clearly divided into two parts as can be seen in *Figure 13.1*; for the first third of the collection the organisation of the two versions is very different, with only two relatively short sequences of recipes occurring in the same order in both versions, while the latter two thirds of the collection are organised relatively similarly in both versions, with longer spans of identically ordered recipes broken by occasional transposed or added recipes. Up to recipe 34 in MS H4016 (corresponding to recipe 20 in MS C), the recipe orders in the two collections have little in common, except for the fact that apart from some individual recipes,<sup>30</sup> this initial part consists of largely

<sup>25</sup> The interpretation that these recipes have been omitted from MS H4016 rather than added to MS C is based on the fact that all of these recipes are also found in the  $\delta$  group, albeit widely scattered in different parts of the collection.

<sup>26</sup> This would mean that recipes PD 346–350, which occur near the lost recipes in MS H4016, did most likely not occur in MS C but are unique to MS H4016.

<sup>27</sup> All of these recipes occur at the point corresponding to the gap in MS H4016, except for recipes PD 69 and 89 which occur together at a slightly earlier point.

<sup>28</sup> Recipes omitted from MS C in its current state but most likely originally included in it have been included in parentheses.

<sup>29</sup> This recipe is not really a genuinely unique one, but rather a conflation of the beginning of recipe PD 171 (“Cabbages”) with a part of recipe PD 172 (“Meat Gele”), which occur consecutively in the  $\delta$  versions but as recipes 8 and 82 in MS C.

<sup>30</sup> These include PD 178 (“Frumenty with Venison”), which has been included with other venison

the same recipes in both versions but in very different order.

After this initial confusion, there is a sequence of 10 co-sequential recipes (from PD 183 to PD 98), broken by the omission of PD 103 (“Cockatrice”) from MS H4016, followed by a sequence of 16 recipes where the only anomalies are the inversion of recipes PD 28 and 109 (for stewed and roasted partridge) and recipes PD 112 and 113 (for roasted egret and curlew) between the two versions. After the sequence of recipes omitted from MS H4016, mentioned above, and the two recipes for chicken (PD 69 & 89) that have been transposed to the lost folia in MS C, the versions again share the next 18 recipes in exactly the same order. While the sequence of shared recipes is seemingly broken at this point, it seems likely that also the recipes currently missing from MS C originally occurred in the same order as in MS H4016. After this second missing folio, the two versions proceed in the same order, with the exception of the above mentioned transposition of PD 178 (recipe 7 in MS H4016) to recipe 78 and PD 43 (recipe 29 in MS H4016) to recipe 126 in MS C, the omission of recipe PD 13 from MS H4016 and the division of recipes PD 141 and 149—each of which provides two alternative ways of preparing the dish—into two separate recipes in MS H4016.

Thus, while the similarity of these two versions is significantly lower than that of the other pairs when measured by the metrics presented above in section 13.1, this is mainly due to the damage sustained by MS C on the one hand, and the differing order in the initial part of the collection on the other. Discounting the damage, the last two thirds of the  $\epsilon$  group are very similar to the other two pairs of versions in its internal consistency.

### 13.3 Semantic organization of the versions

While the preceding sections have examined the differences between the internal organisation of the six versions in mostly quantitative terms and from the point of view of the differences in their formal structure, this section will approach their organisation from a more semantic point of view, trying to uncover the culinary rationale—if any—behind the organisation of the collections. As was mentioned in section 8.2 above, the organisation of recipes within medieval recipe collections is usually not entirely random,<sup>31</sup> but follows some kind of a rationale, the most common organisational principles found in English and Continental recipe collections being: 1) alphabetical, 2) by principal ingredient, 3) by type of dish, and 4) following the order of serving. None of these principles seem to be used consistently in any of the *Potage Dyvers* versions, but at least the second and third ones do seem to have influenced the ordering of the recipes in most of the versions. Thus, while none of the versions—like most English collections—exhibits the kind of detailed and explicit structuring found for example in the *Viandier de Taillevent*<sup>32</sup>, all of

---

recipes in the beginning of MS H4016 but occurs only later in MS C, PD 43 (“Leach Frys”), which has been included with other pastry recipes as recipe 29 of MS H4016 but occurs as recipe 126 in MS C, and the omitted or added recipes listed in Table 13.5.

<sup>31</sup> Although Hieatt (2012) describes the recipe collection found in Oxford, Corpus Christi College MS F 291 as having “no discernable overall rationale and no resemblance to the order of any other collection” (11).

<sup>32</sup> The ‘traditional’ part of the *Viandier* is hierarchically divided into a number of sections based on various principles (Scully 1988: 17–20): 1) Recipes for meat dishes, divided into a) boiled meat dishes,

the versions do contain discernible groupings of recipes defined on the basis of the type of dish, the principal ingredient, or the type of day on which the dishes were intended to be served.

### 13.3.1 MSS Ashmole 1439 and Harley 279

These two versions are the most distinctive of the six in terms of both the extent and ordering of their recipes. In terms of their explicit textual organisation, they consist of an untitled longer ‘main’ part, consisting of the first 152 (MS As) or 153 (MS H279) recipes, followed by three (MS As) or two (MS H279) separately labeled ‘subcollections’: *Leche viaunde/leche vyaundez* (Asrecipes 153–216 and H279 recipes 154–217) *Viaundes ffurnes/Here begynnyth dyuerse bake Metis* (As recipes 217–250 and H279 recipes 218–258), and *Sauces pur diuerse viaundes* (As recipes 251–269). Of the explicitly labeled sections, the section of sauces in MS As and the section on baked foods are true to their labels, the former containing recipes for a variety of hot and cold sauces, both ones intended for specific dishes and more general-purpose ones, and the latter containing recipes for a variety of pies, pastries, and baked meats and fish. The first subsection, whose title essentially means ‘sliced foods’, is slightly less uniform in its contents. The first 18 recipes of this section are indeed for dishes that are explicitly served as slices, but the rest have a more tenuous connection with the title. The next 22 recipes (As recipes 153–192 and H279 recipes 154–193) are for dishes that could be characterised as ‘moulded’ or ‘shaped’ foods which would most likely be sliced at the table when served, except for individual recipes for rice flour (174/175), wafers (176/177) and a syrupy sauce (184/185) which seem to be somewhat out of place in this section. These are followed by a series of recipes for various kinds of toasts, fritters and pasties that could be considered to constitute ‘slices’ only in a very extended sense of the word.

The main part of these versions which accounts for almost two thirds of their recipes—the *Potage Dyvers* proper—consists mostly of different kinds of pottages and boiled or otherwise moist dishes, but does not seem to be internally organised according to any further discernible rationale. In terms of principal ingredients, recipes based on different kinds of meats, birds and fishes, as well as vegetable- and dairy-based dishes occur one after the other with no distinction, although there are some local clusters of dishes based on the same principal ingredient.<sup>33</sup> However, it is difficult to find a sequence of more than three consecutive recipes that would be based on the same category of principal ingredient. While the type of dish seems to be a slightly stronger organising principle than their principal ingredient, it is used quite inconsistently and mainly on the local level, connecting together short sequences of recipes. These kinds of sequences include recipes 1–25 (pottages and boiled dishes of various kinds), 26–33 (sops), 43–5 (purees), 61–68 (meats in broth or sauce), and 108–9 (jellies). After the initial section of recipes for dishes explicitly

b) meats in sauce (broths, stews), and c) roast meats, including i) roasted domestic animals, and ii) roasted fowl, and d) entremets. 2) Recipes for meatless dishes, divided into a) fish and vegetables in sauce (broths, stews), b) sick dishes (with and without meat), and c) fish, divided into i) freshwater fish, ii) round sea-fish, iii) flat sea-fish, and iv) shellfish. 3) Recipes for sauces, divided into a) cold sauces, and b) hot sauces..

<sup>33</sup> Recipes 34–5 are based on pears, 39–40 and 92–3 on shellfish, 41–2 and 81–4 on chicken, 94–6 on tench, 124–7 on different kinds of flowers, and 133–4 on apples.

characterised as pottages—perhaps seen as the most archetypal representatives of the category—recipes for different types of dishes seem to intermingle rather freely, meaning that even the type of dish cannot be considered to provide an overarching organisational principle within the section.

The fourth organisational principle listed above, suggested by Hieatt (1988: 16-7) as the predominant one for English collections, does not seem to go much further in explaining the order of recipes in these versions; while they begin with recipes for various kinds of pottages that can be considered plain ‘everyday’ dishes typical to the beginning of a meal, the end of the main part of the collection also contains several recipes for similar pottages (e.g. recipes 144/5, 145/6, 147/8), and more delicate dishes typical to the end of a meal are spread throughout the collection. Furthermore, unlike some of the other versions of the *PD* family, none of the sections in these two versions would seem to distinguish between dishes suitable for meat and fish days, fish and vegetable recipes suitable for Lent occurring interspersed with meat and dairy ones.

### 13.3.2 MSS Additional 5467 and Douce 55

As was pointed out above, the  $\delta$  group constitutes what could be considered the most prototypical version of the *Potage Dyvers* family of collections, being both more internally consistent than the other pairs and containing the largest number of recipes shared by more than two versions. This should not, however be taken to mean that its internal organisation—which differs significantly even from that of the extent-wise closely related  $\epsilon$  pair—would be any closer to an ‘original’ or a more ‘correct’ one. Like the previous pair, also the  $\delta$  versions consist of an untitled main part, which accounts for half of the total recipes in the collection, and four separately titled subsections. The main part of the collection shows very little discernible organisation; individual recipes and short segments of two or three similar recipes for various kinds of meat, fish, fowl and dairy-based dishes are interspersed with little discernible logic. Neither does the type of dish or method of preparation (boiled, baked, fried, or roasted) or its position in the meal seem to be used as a criterion for grouping recipes.<sup>34</sup>

The latter half of these two versions, on the other hand, is organised into four labeled groups, defined on the basis of either the principal ingredient used (2 and 3), the type of dish described (1) or the appropriate days of serving (4). Although only the beginnings of these subsections are explicitly marked by a heading and the subgroups would thus seem to cover the entire latter half of the collection, looking at the recipes makes it obvious that not all of the recipes following the headings belong to the category indicated by it. The first of these subsections is titled ‘Diverse sauces for diverse viands’, and would seem to cover recipes 92–105 in each of the versions, based on the headings. However, it is obvious that only the first seven recipes following the heading (recipes 92–98) are for sauces, the rest being for a variety of miscellaneous dishes. This group is also shared with the

<sup>34</sup> As an example, recipes 60–80 in both of these versions consist of baked lamprey (60), funnel cakes (61), eggs cooked on a bed of salt (62), a buttered cake (63), strips of bread in milk thickened with egg yolk (64), oysters, rabbit and chicken in gravy or broth (65–9), meat and fruit pies (71–72), roast pike in wine sauce (73), pork and fruit fritters (74), poached pike in sauce (75), veal, kid or chicken stew (76–7), boiled mussels (78–9) and poached sturgeon in sauce (80).

ε group where it occurs untitled near the beginning of the collection, and with MS As, where it forms a part of the above mentioned final section on sauces.<sup>35</sup>

Unlike the first one, the second subsection, titled 'The manner for roasting and saucing diverse viands', does indeed seem to cover the entire span indicated by the headings, beginning at recipe 106 and containing 20 (MS Ad) or 24 (MS D) recipes for roasting (and saucing) various kinds of birds and other kinds of meats. This group is also shared with the ε group, where it occurs near the beginning of the collection without any heading or other paratextual indicator of its status as an independent unit. The difference in the extent of this group between MS Ad and D versions stems from the fact that a recipe for roasting *brew* (i.e. 'whimbrel') has been omitted from MS Ad, and MS D also includes some recipes for roast kid, veal and venison which do not occur in MS Ad.<sup>36</sup> In addition to birds, this sequence also includes two recipes for roast rabbit (115–6 in MS Ad and 116–7 in MS D) and one for glazed pork (123 in MS Ad and 127 in MS D). In the ε group, these are located after this sequence. This group is followed by a similar group of recipes for boiling and/or roasting a variety of fish (recipes 127–47 in MS Ad and 130–50 in MD D), titled—quite descriptively—'The manner for roasting, frying and boiling diverse fish'. This group also occurs as a cohesive group—albeit untitled and with minor variations in order—at the end of both of the ε versions.

The final subsection of the δ versions is the least obvious and would be quite difficult to distinguish without the explicit title, 'Diverse viands for Lent'.<sup>37</sup> Although the heading can be taken to imply that all of the recipes following it would be intended for Lent, a closer examination of the recipes indicates that out of the 34 recipes, only the first 22 are suitable for Lent, or at least for fish days. Compared to the three other subsections, the last one seems to be the least well established; while all but two of the recipes included in it occur also in the ε versions and half of them occur also in the γ versions, they do not occur as a coherent group in these versions, but are rather scattered throughout the collection. This would seem to point to the section of Lenten recipes being an innovation of MS δ, being either imperfectly realised in the first place or adulterated with non-Lenten recipes upon subsequent copying.

Looking at the occurrence of these consistent groups of recipes in the different groups of versions, we can see that while the majority of the miscellaneous recipes are shared between all of the versions, none of the recipes from the subsections of fowl and fish recipes shared between the β versions occur in the γ versions. This means that these groups of recipes were either consistently omitted from MS γ (or its ancestor) or not included in any ancestor shared by all of the surviving versions (represented by α) and added to MS β from another source. It is of course impossible to determine this with certainty, but the fact that these recipes are preserved as a unified whole in all of the four β versions, along with the lack of any indication that they would have been deliberately omitted from the γ versions (e.g. the complete lack of fish and fowl recipes in these versions, or the inclusion of alternative recipes for fish and fowl from some other source), would

<sup>35</sup> The seven recipes that are not for sauces also occur in the ε group, although scattered around the collection. Three of the recipes, 103–5, also occur in the γ group, recipes 103–4 occurring as a pair.

<sup>36</sup> These recipes do, however occur in the ε group, although in a different place after the bird recipes.

<sup>37</sup> The title in MS D actually reads 'Diverse *sauces* for Lent', which is obviously a mistake, most likely caused by interference from the title of the first subsection.

seem to suggest that they have been added to an ancestor shared by MSS Ad, D, C and H4016 but not by MSS As and H279, represented by MS  $\beta$  in *Figure 13.2*.

### 13.3.3 MSS Cosin V.iii.11 A and Harley 4016

Although no groupings of recipes are explicitly indicated in either of these manuscripts, several levels of organisation can nevertheless be distinguished. The most fundamental of these is a quite clear division into meat and fish day dishes (most of the latter also being suitable for Lent): those up to recipe 82 in MS C and 94 in MS H4016 are principally meat dishes,<sup>38</sup> This division would also seem to agree with the mutual relationship observed above for these two versions; while the order of recipes in the initial part of the collection is quite different in the two versions, there is only one recipe that crosses the boundary mentioned above, occurring in the initial part of one version and the final one of the other.<sup>39</sup> Although the initial part of the collection in both versions contains some recipes that are suitable also for fish days,<sup>40</sup> only the pea puree in recipe 72 (MS C or 81 (H4016) is a dish that would be likely to be associated primarily with fish days. All of the recipes from recipe 83/95 onwards are for meatless dishes, the great majority (73 out of 86 in MS C) being also suitable for Lenten diet.

In addition to this broad division, there are also several groupings of recipes, based on both the principal ingredient of the dishes and their method of preparation. These groupings do not, however, cover all of the collection in either version, and there are also some recipes whose location is not explicable in terms of them. Both of the versions open with a group of boiled meat dishes (recipes 1–13 and 15–16 in MS C and 1–19 in MS H4016), mostly consisting of various types of pottages, many also incorporating vegetables.<sup>41</sup> The next group seems to consist of various kinds of pastries (recipes 14, 17–19 and the missing recipes in MS C and recipes 20–33 in MS H4016).<sup>42</sup> These are followed by the group of sauces shared with the  $\delta$  group (recipes 21–29 in MS C and 35–43 in MS H4016). Next comes the group of roasted birds occurring as a separate subsection in the  $\delta$  versions (recipes 31–46 in MS C and 44–59 in MS H4016), which has here been extended by including a recipe for boiled partridge (34/48) (found elsewhere in the  $\delta$  versions and also in in MSS As and H279) and following it with a group of recipes for chicken and rabbit, as well as a seemingly misplaced recipe for stewed beef (MS C 49).<sup>43</sup> The next discernible group (recipes 53–69 in MS C and 60–78 in MS H4016) consists of

<sup>38</sup> The conformance of the recipes missing from MS C to this pattern is uncertain, but judging from the recipe titles given in the table of contents and the recipes found in corresponding locations in MS H4016, they would seem to describe either meat dishes or dishes suitable for both fish and meat days.

<sup>39</sup> This exception is recipe PD 43, which occurs in the beginning of MS H4016 and near the end of MS C.

<sup>40</sup> These include the group of sauces shared with the  $\delta$  versions (recipes 23–29 in MS C and 37–43 in MS H4016), and such mainly egg-based dishes as fritters (recipe 14/23), custard tarts (17/20), french toast (70-1/79–80), pea puree (72/81), scrambled egg toast (73/82) and a herb omelet (80/92)

<sup>41</sup> These dishes are exactly the kinds that were supposed to be served as the first substantial course of a feast (see subsections 7.1.1 and 7.2.3), lending credence to the hypothesis of Hieatt (1988: 16–7) about the order of serving being used as an organisational principle, if not *the* organisational principle of Middle English recipe collections.

<sup>42</sup> In the table of contents of MS C, recipe 14 ('Long Fritters') is located after the boiled meat dishes and just before recipe 17, making it the first recipe of the pastry group.

<sup>43</sup> As noted above, MS H4016 entirely omits 6 recipes found in MS C at this point and also found in

5 recipes for coney and rabbit followed by three recipes for roasting kid, venison and veal and a group of recipes for stuffed and endored meats, followed by four recipes for various meat dishes, all characterised by more than one stage of preparation and an end result that is more complex than a simple roast or pottage. The meat day section is closed by a selection of rather miscellaneous recipes (70–82, along with the recipes lost with the missing folio in MS C and recipes 79–94 in MS H4016), several of which are also suitable for fish days (although not for Lent), and may be located here simply because the compiler found no better place for them.

The meat and fish day dishes are divided by a group of recipes for various sweet dishes of cooked fruit (recipes 84–87 in MS C and 96–99 in MS H4016), which could be considered to belong to either category, since fruit cooked in wine were a common ‘dessert’ and their eating was not restricted by the rules of abstinence. Considering that the collection began with a selection of recipes suitable for the beginning of a meal, it would be tempting to apply the rationale suggested by Hieatt (1988) and to consider these recipes as the end of the meat section rather than the beginning of the fish section. The first proper fish day group consist of a variety of fish and vegetable soups and stews (88–100/100–12), including a recipe for almond milk and a fish day mortreus based on it. Next come three milk-based dishes, suitable for fish days but not for Lent. They are followed by a selection of rather miscellaneous fish day recipes (104–117/115–28), paralleling the group of similar meat dishes above.<sup>44</sup> After these miscellaneous recipes, the first distinguishable group is formed by the four soups or broths (recipes 118–21 in MS C and 129–32 in MS H4016), including also another recipe for almond milk. These are followed by a selection of sweet fritters, pastries and cakes (122–30/133–140), all either implicitly or explicitly intended for fish days or Lent.<sup>45</sup>

The final part of the collection—save for the very last recipe in MS C and the last two recipes in MS H4016—is made up of the section of fish recipes shared with the  $\delta$  group which forms the largest (38 recipes) cohesive group in the collection.<sup>46</sup> Unlike in the  $\delta$  group, the fish recipes here seem to be grouped according to species, although no distinction is made between sea and freshwater fish. The  $\epsilon$  versions also incorporate into this sequence a further recipe for ‘Sturgeon in Broth’ that occurs elsewhere in all of the other versions, suggesting a conscious reorganisation of these versions in comparison to the  $\delta$  versions. In both of the versions the collection is concluded, rather strangely, by a recipe for a pottage of milk, flour and eggs that is titled ‘Papins’ and found in the beginning of the collection in all but these versions. After this recipe which already seems like an afterthought or something missed in copying, MS H4016 adds a further unique

---

the  $\delta$  group elsewhere in the collection.

<sup>44</sup> These miscellaneous recipes, of which some are also suitable for Lent, include recipes for the popular ‘Leach Lombard’ or ‘Lombard slices’ of dried fruit (104–5/115–6), fritters (106/117), eggs cooked on a bed of salt (107/118), fruit pastries fried in oil (108/119), poached eggs (109/120), a thick pottage of almond milk and fruit (110/121), puddings of dried fruit (111/122, 113/124 and 115/126), pottages of roe and fish intestines (112/123, 114/125 and 116/127) and fish jelly (117/128).

<sup>45</sup> Only two of these recipes, PD 43—which is recipe 126 in MS C but occurs in the meat-day section of the collection as recipe 26 in MS H4016—and PD 63, which is recipe 128 in MS C and recipe 138 in MS H4016, are not suited for Lent, the former using cheese and the latter butter.

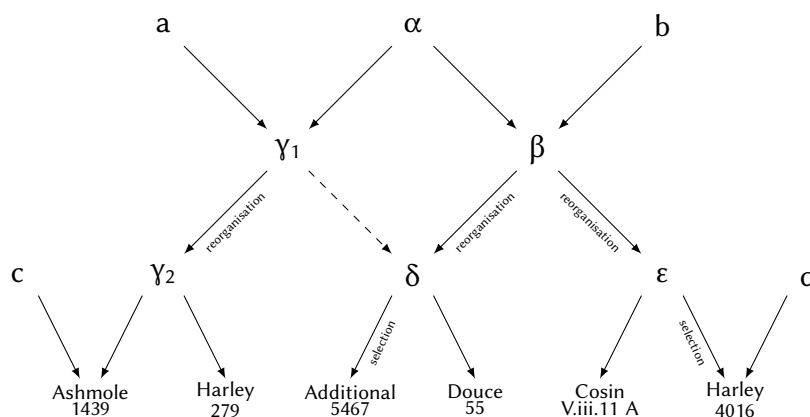
<sup>46</sup> In MS H4016, this section contains 40 recipes, as the recipes for ‘Boiled or Roast Gurnard’ (PD 141) and ‘Tench in Brasee’ (PD 149), each of which provide two alternative ways of preparing the fish, have been divided into two separate recipes in MS H4016.



recipe—most likely from a completely different source—for ‘Quince Paste’ which is quasi-medicinal in nature, containing unusually precise measurements for spices and being mentioned as being “comfortable for a mannys body And namely fore the Stomak”.

## 13.4 Conclusion

As the macro-level structural analysis undertaken in this chapter has shown, the *Potage Dyvers* family provides an excellent demonstration of the utility of structural analysis in elucidating both the multiple textual history of the work and any semantic or functional principles underlying the organisation of the discourse colony. The analysis shows a strongly bifurcating tendency in the transmission history of the structure of the discourse colony, as the relationships between the organisation of the different versions are characterised either by a very high or very low degree of similarity. Already a simple quantitative analysis of the extent and ordering of the recipes contained in the six manuscript versions of the *Potage Dyvers* revealed clear similarities and differences among the versions and allowed the formulation of an initial hypothesis about their familial relationships. A closer analysis of the spatial distribution and qualitative nature of these differences revealed further details about the likely histories of distinct portions of the collection and the likely points in the history of the different versions of the work where additional material was added. Based on the above structural analysis, the concept of *family* would indeed seem to be an accurate characterisation of the relationships between the six manuscript versions; not only are the different surviving versions characterisable as ‘siblings’, ‘cousins’ and ‘second cousins’ to each other, but the apparent inclusion of material from several sources into the different versions also reinforces the semblance of a bilineal family tree.



**Figure 13.4:** The transmissional relationships of the six surviving *Potage Dyvers* versions and their postulated ancestors ( $\alpha$ – $\epsilon$ ,  $a$ – $d$ ) suggested by the analysis of the extent and ordering of recipes contained in them.

Figure 13.4 contains a diagrammatic representation of the transmissional relationships between the six surviving versions and their hypothetical ancestors based on their textual structure, as suggested by the analysis undertaken in this chapter. In the diagram, Greek letters ( $\alpha$ – $\epsilon$ ) represent the different transmissional stages postulated for what has here been called the *Potage Dyvers* family,  $\alpha$  representing the ancestor shared by all of the surviving versions that could be considered the ‘*ur-version*’ of this family, while the Latin letters (a–d) represent other sources of recipes, considered in some respect ‘external’ influences on the tradition. The arrows represent the transmission of recipe material through copying, and the labels next to some of the arrows indicate alterations in the extent or ordering of the recipes. As has become clear above, the six surviving manuscript versions of the *Potage Dyvers* analysed here are clearly divided into three closely related pairs that share a common core of recipes, most likely inherited from a shared ancestor. In addition to this common core, which accounts for roughly half of the total in each pair (one third in the case of the  $\gamma$  pair), all of them also include recipes that do not seem to be derived from the common ancestor, as they are not shared by all of the pairs. In this respect the family seems to be divided into two groups, as shown in Figure 13.4: the  $\gamma$  pair, consisting of MSS As and H279 contains a large proportion of recipes that are not shared by the other four versions, being most likely copied to an ancestor shared by these two versions ( $\gamma_1$ ) from a separate source (a),<sup>47</sup> while the  $\beta$  group, made up of the other four versions, includes a considerable number of recipes that are not shared by the  $\gamma$  pair, being most likely copied from another source (b) to an ancestor shared by all of them ( $\beta$ ). These two groups would seem to account for the majority of the genuine differences in extent between the versions (i.e. not resulting from later damage), and the family as a whole would then seem to have three principal sources:  $\alpha$ , a, and b.<sup>48</sup>

Based on the ordering of the recipes in the surviving versions, it would seem that both the  $\gamma_1$  and the  $\beta$  versions were at some point of their transmission reorganised to produce versions  $\gamma_2$ ,  $\delta$ , and  $\epsilon$ . In the case of  $\gamma_1$  and  $\gamma_2$ , this hypothesis is based solely on the fact that while the ordering of the recipes is practically identical in MSS As and H279, the recipes unique to them are scattered throughout the collection, which suggests that the  $\gamma$  version of the collection was reorganised at some point after the copying of the recipes from a to integrate them to the recipes inherited from  $\alpha$ .<sup>49</sup> In the case of  $\beta$ , this reorganisation is indicated by the fact that while the  $\delta$  and  $\epsilon$  pairs share the vast majority of their recipes, they are presented in very different order between the two pairs. While the  $\delta$  pair is organised into

<sup>47</sup> In the case of the  $\gamma$  group, the addition of material from two distinct sources is also indicated by the occurrence of two linguistically distinct versions of the same recipe in both of the  $\gamma$  versions, which would be very unlikely in a case of strictly unilineal transmission.

<sup>48</sup> As was noted in subsection 9.3.1, the b source would seem to be somehow related to the families of collections known as *An Ordinance of Pottage* and *A Noble Boke off Cookry*, as the  $\beta$  versions contain almost 30 recipes contained also in these families, among them the sequence of recipes for birds mentioned above. The fact that these two families also contain some recipes here associated with  $\alpha$  points towards the existence of an extensive and complicated network of connections between the 15<sup>th</sup>-century English families of recipe collections.

<sup>49</sup> It is of course possible that this integration was done already at the point of copying, the scribe alternately copying recipes from the two sources, but in order to emphasise this operation of integration on a conceptual level, it has here been described as a separate stage.

a general main section and four thematic subgroups—similar but different from those in the  $\gamma$  group—the  $\epsilon$  pair seems to be organised into separate, although unlabelled, sections of meat-day and fish-day recipes.<sup>50</sup> Although there is little evidence, it the  $\delta$  group whose organisation would seem to more closely approximate the organisation of the hypothetical  $\beta$ , based not only on the homogeneity of the two  $\delta$  versions but also on the evidence provided by recipe PD 343 occurring in MS C, which conflates together parts of recipes PD 171 and 172 which occur widely separated in this version but consecutively in the  $\delta$  versions.<sup>51</sup>

Although the three pairs are remarkably consistent internally—apart from the initial part of MSS C and H4016 which has been reorganised in one of the collections—some modifications to the extent of the surviving versions have also occurred after the initial division into the  $\gamma$  and  $\beta$  groups. On the level of the three sibling pairs, there seems to be very little interaction between the two main branches, although there is some evidence of either ‘cross-contamination’ from  $\gamma_1$  to  $\delta$  or from of a shared external source between these two versions—marked with a dotted line in *Figure 13.4*—in the form of three recipes, PD 50, 70 and 177 (of which the first does not occur in MS As), that are shared between the  $\gamma$  and  $\delta$  groups but do not occur in either of the  $\epsilon$  versions.<sup>52</sup> The slight differences in extent between the sibling MS versions also imply the addition of some recipes in the last stages of their textual transmission. For example the unique section of sauces appended at the end of MS As clearly indicates one or more external sources beyond  $\gamma_2$ . The fact that some but not all of the sauce recipes also occur in the  $\beta$  group may be taken as an indication that the source c was related to b but also incorporated material from other sources. The differences in the extents of MSS Ad and D would seem to be entirely the result of the omission of 4 recipes from MS Ad, while the differences between the  $\epsilon$  pair are indicative of a more complicated situation.

Of the  $\epsilon$  pair, MS C seems to be the more ‘conservative’ one, the majority of the changes in extent occurring in the copying of MS H4016.<sup>53</sup> These changes involve the omission of a number of recipes occurring in the rest of the  $\beta$  manuscripts, the most notable of which is a group of recipes that occurs as consecutive series in MS C (PD 27, PD 29, PD 3, PD 129, PD 102, and PD 35) between recipes 46 and 53, which correspond to recipes 59 and 60 in MS H4016, the most likely explanation for which is a damaged exemplar occurring somewhere between  $\epsilon$  and MS H4016.

<sup>50</sup> The reason why the two major groups of recipes that are here postulated to have been added to  $\beta$  from the external source labelled b have been preserved more or less intact most likely has to do with their strong internal coherence in terms of their semantic content.

<sup>51</sup> This would seem to imply not only that the order of the  $\delta$  group is closer to the ‘original’ one, but also that the conflation of these two recipes was not caused by the MS C scribe but happened before the recipes were reorganised for MSS C and H4016 but most likely after the  $\delta$  and  $\epsilon$  groups diverged from each other. The fact that both of the recipes are also included individually in all of the six PD versions edited here makes it likely the scribe who conflated them—most likely by accident, beginning the first one and then jumping into the second midway—realised his error and copied also the intact versions. Since it nevertheless occurs in MS C, this conflated recipe—whether marked for deletion or not—was passed on by the scribes of (the hypothetical ancestor) version  $\epsilon$  and of MS C, but apparently not by the scribe of MS H4016 or one of its unique ancestors.

<sup>52</sup> It is, however also possible, that these recipes are a part of the common core but were either accidentally or purposefully omitted from  $\epsilon$ .

<sup>53</sup> The only differences that are not known with certainty to be due to damage to MS C but would seem to involve its copying would seem to be the omission of PD 22, PD 40, and PD 58, all of which occur in all of the other versions (and could equally well have occurred also in the missing sections of MS C), although they do not seem to be mentioned in the table of contents.

The other recipes missing from MS H4016 (PD 1, PD 33, PD 85, PD 99, PD 103) are all shared by all of the other versions, which makes it likely that they were part of the shared  $\alpha$  version and have been either accidentally or purposefully omitted from MS H4016 or its post- $\epsilon$  ancestor. In addition to the omission of these recipes, MS H4016 also includes some unique recipes (PD 344–53), occurring mostly in its initial part and most likely copied from an external source, marked as d in *Figure 13.4*.<sup>54</sup>

Based on the above, the characterisation by Hieatt (2004: 27) of MSS H279 and As as “the same collection” and MS H4016 as a different collection but “of the same ‘family’” would seem to be accurate. Whether we wish to consider these six versions to constitute two or three separate works of the same family depends mostly on the relative emphasis we place on the *extent* of the versions on the one hand and on their *ordering* on the other. If we take into account just the extent and maintain the relatively loose tolerance for variation mentioned in subsection 2.1.4, namely that two *versions* of a discourse colony that represent the same work contain more shared than unique material, the most obvious division is into two separate works, one surviving in two manuscript versions and the other in four.<sup>55</sup> If, on the other hand, we consider also the *ordering* of the collection to form a part of its identity as a discourse colony, we end up with a family of at least three separate works—possibly even four or five, if we consider the differences in the extent of MSS As and H279 and in the ordering of MSS C and H4016 to be sufficiently substantive. While the ordering of the component parts in a discourse colony does not by definition affect the meaning of the individual components, I would argue that its organisation nevertheless has sufficiently significant pragmatic function to be considered a significant part of its *identity* as a work. Since different levels of distinction are useful in different situations and for different purposes, it might be useful to define a further categorisation into *sub-families* below the level of family. This would allow us to see the *Potage Dyvers* family as consisting of two subfamilies, one containing two different works and the other three, as shown in *Table 13.5*.<sup>56</sup>

As to the clues this analysis provides about the copying practices applied to medieval recipe collections—and possibly also to other kinds of discourse colonies—the most significant is the affirmation of the considerable freedom with which the scribes seem to approach these texts. The copying practices employed by the scribes in copying these kinds of texts seem to be strongly polarised between complete faithfulness and comprehensive reorganisation, resulting in copies that are either practically identical or significantly differ in their organisation. There do not seem to be any obvious groups of recipes that would maintain their structural

<sup>54</sup> The individual recipe appended to the end of MS H4016 has most likely been copied separately from some other source.

<sup>55</sup> Taking a stricter approach to variation in extent but ignoring the ordering of the versions, on the other hand, would result in a distinction between four separate works, realised by 1) MS As, 2) MS H279, 3) MSS Ad, D and C, and 4) MS H4016, although this division would intuitively seem to place too much emphasis on the extent of the collections at the cost of their ordering.

<sup>56</sup> Since the extent and ordering of the components parts is only one aspect of the similarity of discourse colonies, the other being the internal similarity of those component parts, the division of the versions into separate *works* on this criteria alone is slightly problematic, and an analysis of the internal similarity of the individual recipes might give cause for a re-evaluation the of these divisions in either direction.

family	sub-family	work	version
Potage Dyvers	PD A	PD A1	PD MS As
		PD A2	PD MS H279
	PD B	PD B1	PD MS Ad
			PD MS D
		PD B2	PD MS C
		PD B3	PD MS H4016

**Figure 13.5:** *The division of the Potage Dyvers family into subfamilies, works and versions based on the extent and ordering of recipes.*

integrity in all of the surviving versions beyond individual pairs of recipes. Even within the smaller groups and pairs of versions, the order of recipes is maintained either more or less globally or not at all, there being very few recipe sequences that would occur as a unit in an otherwise heterogeneously organised context. This freedom seems to have applied not only to the throughgoing reorganisation of the material of a single exemplar according to a different paradigm, as seen in the  $\beta$  group, but also to the combination of material from different sources representing what could be considered entirely different collections. The medieval producers and users of recipe collections do not seem to have viewed these families of collections as *works* but rather as relatively amorphous repositories of useful knowledge which could be either copied as they were if they seemed useful as such, or mined for groups of recipes to be combined with ones from other sources to compile a new discourse colony for the specific needs of a new user.

From this point of view it becomes purely an arbitrary modern convenience to talk of the ‘*Potage Dyvers* family’ of collections or to treat it as if it were a group of related *works*. This amorphous and ‘unbounded’ nature of the *Potage Dyvers* texts and other medieval recipe collections, both on the level of their textual tradition and on the level of the selection of recipes contained by individual manuscript versions also means that it is in the end a misguided attempt to consider the surviving versions as entities and assign them to ‘families’ or other types of categories. Rather than seeing the manuscript versions as a whole belonging to a given family or group, it might be more fruitful—although in many cases undoubtedly also significantly more difficult—to try and distinguish groups of recipes with shared textual ancestries and to characterise surviving manuscript versions in terms of the groups from which they contain recipes, in order to eventually build a map of how these groups have travelled and been joined together with other groups to make the collections that have survived to us.

Based on the internal structure of the surviving members of the *PD* family, we can distinguish five groups of recipes with distinct sources and textual histories, represented by  $\alpha$ , a, b, c, and d in *Figure 13.4*. The recipes included in each group are listed in *Table 13.6*, while *Figure 13.6* visually represents the distributions of these groups in the six *PD* versions. In terms of the sources from which they draw their recipes, the six versions could thus be characterised as follows, with minor sources given in parentheses:

MS Ashmole 1439	$\alpha$ -a-(c)
MS Harley 279	$\alpha$ -a
MS Additional 5467	$\alpha$ -b-(a)
MS Douce 55	$\alpha$ -b-(a)
MS Cosin V.iii.11 A	$\alpha$ -b
MS Harley 4016	$\alpha$ -b-(d)

From this point of view the ‘family’ of collections opens up into a concept referring to collections that share recipes with each other, its boundaries determined not absolutely but contextually in relation to a specific text chosen as a reference point, much like in genealogy. This opening up of the family tree would make the collections that are not traditionally considered as a part of the *Potage Dyvers* family but share material with the collections edited here (see section 9.3) as much a part of the family as the six manuscripts edited here. This means that one of the natural continuations for the work undertaken here would be 1) to extend the analysis first to these known relatives, and 2) to edit further collections following the same principles used here, so that algorithmic means of detecting textual similarity—currently used for example for detecting plagiarism—could be used to discover previously undiscovered parallel versions of recipes in more distantly related collections.

Considering the fact—demonstrated by the above analysis—that at least manuscript recipe collections and most likely other kinds of manuscript discourse colonies—and perhaps even many ‘mainstream’ manuscript texts—are not transmitted unilineally but combine content and features from several different ‘blood-lines’, this kind of a multilineal genealogical model might prove to be quite useful for the description of the macro-level structural relationships between medieval manuscript texts. On the micro-level of the textual and linguistic structure of individual recipes or non-colony texts this kind of an approach could well prove to be impossible due to the drastically increased complexity of natural language in comparison with the relatively simple recipe sequences discussed here, but the practical evidence found here of the multilineality and ‘cross-contaminating’ nature of medieval text transmission could well prove useful in the development of new algorithmic methods in the cladistic analysis of manuscript stemmatics.

Thus another natural step in the analysis of the textual transmission of these kinds of discourse colonies would be to extend the analysis to within individual recipes, examining the ways in which they have been syntactically and text-structurally reorganised in different versions in order to trace their individual development as they move from one collection to the next. Since it is—at least in the case of the *PD* family—“usually impossible to say whether a particular version of a recipe is closer to one or another of the parallels” (Hieatt 2004: 31), this would require not only the word-level collation of all parallel versions of the recipes—preferably also those found in manuscript versions which are not closely related on the level of the whole collection—but also the application of *cladistic* analysis<sup>57</sup> separately to each recipe and the aggregation of the data obtained, and was

<sup>57</sup> *Cladistic analysis* is “a technique developed in evolutionary biology to reconstruct the ‘family tree’ of related species by study of the characteristics they share and do not share”, whose applications to manuscript stemmatology have been studied already since the 1990s (Robinson 1994: 89; Robinson 1998: 257). Although the parallels between cladistics and stemmatology have been known for long,

Group $\alpha$ (89)	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 27, 28, 29, 33, 34, 35, 36, 37, 38, 39, 40, 42, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 68, 69, 71, 72, 74, 80, 81, 82, 83, 84, 85, 87, 90, 91, 99, 103, 104, 105, 151, 152, 154, 155, 157, 158, 160, 164, 165, 166, 169, 171, 172, 173, 174, 180, 183
Group $a$ (164)	186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 365, 366, 367, 368, 369, 370, 371
Group $b$ (92)	9, 19, 25, 26, 30, 31 <sup>*</sup> , 32, 41 <sup>*</sup> , 43, 44, 66, 67, 73, 75, 76, 77, 78, 79, 86, 88, 89, 92 <sup>*</sup> , 93, 94 <sup>*</sup> , 95 <sup>*</sup> , 96 <sup>*</sup> , 97 <sup>*</sup> , 98 <sup>*</sup> , 100, 101, 102, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 153, 156, 159, 161, 162, 163, 167, 168, 170, 175, 176, 178, 179, 181, 182, 184
Group $c$ (19)	31 <sup>*</sup> , 41 <sup>*</sup> , 92 <sup>*</sup> , 94 <sup>*</sup> , 95 <sup>*</sup> , 96 <sup>*</sup> , 97 <sup>*</sup> , 98 <sup>*</sup> , 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364
Group $d$ (10)	344, 345, 346, 347, 348, 349, 350, 351, 352, 353
Indeterminate (5)	50, 70, 177, 185, 343

<sup>\*</sup>) Shared by groups  $b$  and  $c$ .

Table 13.6: Recipes of the six Potage Dyvers versions (identified by their PD recipe reference number) derived from the different sources identified in Figure 13.4, complete with the total count of recipes from each source.

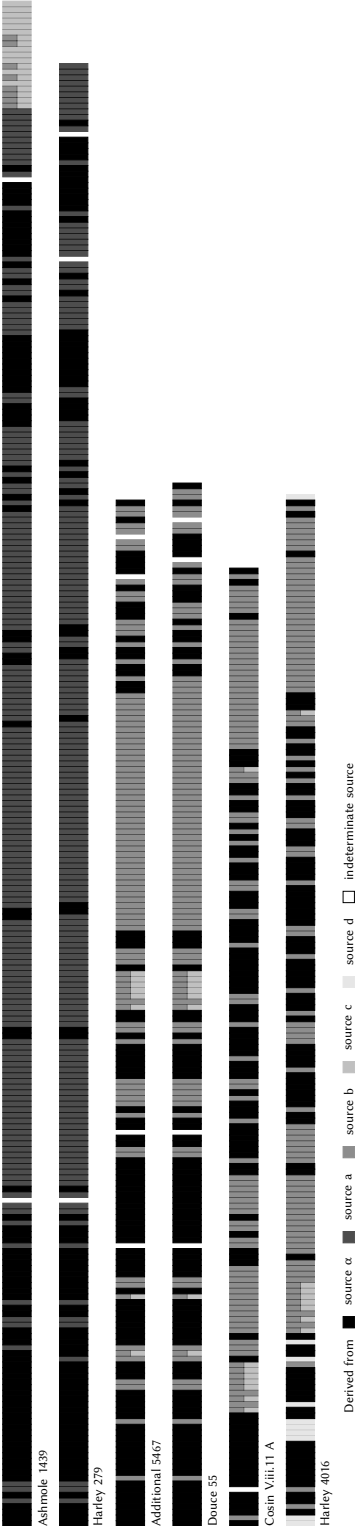


Figure 13.6: The distribution of recipes from different sources in the six Potage Dyvers versions.

thus not possible within the scope of this thesis. However, since computerised structural analyses based on cladistics have been found to be extremely useful in making comprehensible textual traditions that otherwise defy clear exposition (Robinson 1994: 87), and methods of cladistic analysis designed specifically for manuscript stemmatics have recently been developed (see e.g. Roos, Heikkilä and Myllymäki 2006 and Roos and Heikkilä 2009), an exploration of such methods will provide a suitable goal for a separate project, accompanied by the extension of the annotation scheme described in chapter 11 to cover the detailed stand-off annotation of collation results.<sup>58</sup>

---

the earliest application of this method to manuscript studies that Robinson knows of was in an unpublished paper presented by Arthur Lee to the 1987 Patristics conference in Oxford.

<sup>58</sup> Fortunately, the *CollateX* tool (<http://collatex.net/>) developed by the Interedition project (<http://www.interedition.eu/>) would seem to provide the basic functionality required for this operation in the form of a Java-based tool that can be integrated into a Web-based editorial toolset.



## Chapter 14

# Conclusion

We have ample indications of the strengths and weaknesses of traditional methods and theories; it would be valuable to see the capabilities of other kinds of editions. Indeed, the longer the production of humanist editions prefigures the discursive field, the less likely Middle English textual criticism will ever expose and get beyond their theoretical horizon.  
(Machan 1994: 192-193)

In defining and implementing a specific kind of digital diplomatic edition designed to meet the needs of historical corpus linguistics, this thesis and the included edition are intended as the kind of examination of the “capabilities of other kinds of editions” that Machan called for twenty years ago. While it is by no means unique in this sense, it is still much more of a rarity than would be ideal. The fact that over fifteen years after Machan’s statement and twenty years after the first serious experiments at digital editing, Hajo (2010), writing in the context of documentary editing, could still observe that few digital editors so far have “really explored or thought through” the influence of digital editing on “editing’s guiding premises” and that “most digital editions still closely resemble the books they are based upon”, is a stark reminder of the relative paucity of digital editions that challenge the traditional editorial paradigms described in chapter 3.

This is both surprising and somewhat disappointing considering the considerable amount of discussion digital editing has prompted in the secondary literature. As Modiano, Searle and Shillingsburg (2004) observed already ten years ago, the new digital medium has succeeded in replacing old debates about the value of the author’s early and late intentions and the role of editors and other “outsiders” by new ones about issues like “how to present the multiplicity and richness of variant texts, or how to represent agency in these processes, or how to construct electronic navigational aids to enable a sense of immediacy for users confronting a multiplicity of variant texts without losing their way, or how to minimize or eliminate the effects of editorial biases and misunderstandings” (xiii). Although this discussion in itself can be considered a significant achievement, the benefits of

digital editions have more often than not been expressed in terms of their *potential* and discussed in the future tense.

To move the discussion of the possibilities of digital editing away from the potential future and towards a more real present, this thesis has presented the reader with not just a discussion of the benefits of a properly digital editorial approach for the specific field of historical corpus linguistics (chapter 4), but also a set of concrete specifications of how this kind of an edition can be prepared (chapter 10) and encoded (chapter 11). Perhaps even more importantly, it demonstrates what this kind of an edition could look like as a *data archive* (appendix A) and how this data archive can be presented in different ways (appendices B, C and D) and used for different kinds of analyses (chapters 12 and 13).

One reason for the relative paucity of digital editions could be that the theoretical and technical solutions of digital editions are still seen to be in a state of flux and “characterized more by innovation, experimentation, and new developments than by established practices” (Deegan 2006: 362-3), leading to a perception of “the stable or stabilized paper text” (Sutherland 2009: 22) still being the safest option in terms of longevity. This may very well be the case for editions of originally printed works which can be reasonably accurately reproduced in print, but as was pointed out in chapters 3 and 4, the ‘stable printed text’ is simply an anachronistic misrepresentation in the case of medieval vernacular manuscript texts, and what we need instead is a way of “enabling the recovery of the variable text” (Edwards 2000: 78), for which the digital medium is here argued to provide the best solution.

However, as Johanna Drucker recently argued in the *Los Angeles Review of Books*, despite the valuable work done by the Text Encoding Initiative and the various Digital Humanities organisations, the public discussion about the digital medium in humanities scholarship—and especially the promotion of its virtues—has unduly focused on the *presentation* aspects of digital publications—the ‘bells-and-whistles’ of digital publication—to the detriment of the possibilities it opens up for the structuring of scholarly content:

The really innovative aspects of digital work are under the hood and in the nitty-gritty of what we call ‘back-end’ work as much as in the front-end user experience, the multiple pathways or complicated routes through a site or collection of materials. In fact the most exciting and innovative aspects of digital presentation are the ways in which structured data—texts with humanly embedded organization—can be searched and analyzed, not the way they are presented.

(Drucker 2014)

With its focus on the digital edition as a *data archive* instead of an editorial presentation of this data and on the development of processes and formats for the *production* and *encoding* of scholarly research data instead of its presentation, this thesis very much shares this view. While the development of presentation and analysis tools for digital humanities research data—like text editions and linguistic corpora—is vital for our ability to make full use of the data we produce, the argument made here is that it is best undertaken as a separate endeavour, independent of the production of any specific set of data and based on well-documented data formats and interfaces. As was argued in chapter 4, we need to take very seriously the requirement of separating our data from the means of its presentation,

and allow the development cycles of the two proceed independently at their own, vastly different, paces.

## 14.1 Corpus-linguistic editions

Considering that the intention of this thesis was to formulate an editorial praxis for producing editions of historical documents that are usable as corpus texts—complete with the required metadata—it is prudent to briefly return to the issues of editing for historical corpus linguistics mentioned in the Introduction and briefly examine how they have been solved by the present edition and the principles and practices described in chapters 10 and 11.

Traditional linguistic corpora are problematic for pragmatically oriented historical corpus linguistics because they abstract the text away from its original documentary context and eliminate most of the material paratext that can not only serve various pragmatic functions but also help to explain linguistic or textual features of the text. In the editorial approach described here this problem is ameliorated through the detailed *descriptive annotation* of a variety of visual and physical documentary features on a level of abstraction corresponding to the graphemic transcription of the textual content.<sup>1</sup> This allows these features—e.g. the relative position of a textual unit on the manuscript page, its visual distinctiveness (size, colour, shape) in relation to the surrounding text, and its genetic status (i.e. whether they were a part of the initial writing act or of a subsequent emendation)—to be used as search criteria alongside the graphemic identity of textual items. This makes it possible to examine things like the correlation of scribal errors with specific linguistic constructions, the correlation of abbreviation with the position of the item on the page, or the relationship between visual and grammatical highlighting, as well as to relate the use of these features to specific text types or genres, historical periods or even specific scribal styles. Since all of this annotation is encoded as separate XML elements, it can be hidden or ignored when it is not relevant and therefore does not hinder the use of the corpus in any way.

Most existing digital editions that have been published either online or on physical media like a CD-ROM are presented through a specific user interface or in a format intended for reading or visual perusal of the text. Even if search functionality is provided, it rarely meets the needs of corpus linguists, who would usually require at least a concordance search with exportable KWIC results, ideally with a persistent link between the individual concordance lines and the corpus. The most obvious solution for circumventing the limitations of the user interface would be to extract the texts from the edition and use external tools to analyse them, but very few digital editions allow the extraction of the edited texts—much less any metadata—from the edition in a format that would allow them to be used as a part of a corpus. The solution proposed here is to conceive the core of the

<sup>1</sup> It could be argued that facsimile images of the original texts—traditionally used by pragmatically oriented corpus linguists to qualitatively relate their corpus findings to the original document context—could also serve this purpose. However, even when they are available, they have the disadvantage of not being searchable, which limits their use to qualitative analysis separately from any quantitative analyses of the textual content. For this reason they have not been considered to be an essential part of a corpus linguistic edition, but rather ‘something extra’, which can add value to the edition but cannot replace the descriptive annotation of documentary features.

edition as a *data archive*, independent of any user interfaces, tools or editorial presentations, and always make this data archive—containing all of the data and metadata—available to the user in a standardised format, in this case as a TEI XML document. Any presentation formats (such as the diplomatic and parallel reading versions of the present edition) are not considered to be alternative ‘forms’ of the edition but rather derivatives of it, and any user interfaces used to present it are similarly considered merely means of accessing, presenting, or analysing the edition, not an integral part of it.

The problems caused by spelling variation in the quantitative lexical analysis of historical corpora are well-known. Not only is it difficult to find all forms of a certain lexical item, but the high degree of variation also compounds the problem of homonymy, as the multiple variant spellings of different Middle English—and even Early Modern English—words overlap with each other to a much greater extent than in Present-Day English. This problem has sometimes been solved through the wholesale lexical normalisation of the corpus, but this has the downside of precluding all research questions that are interested in the dialectal or other orthographic variation in the text. While the provision of both normalised and original versions of the corpus texts does cater to a wider variety of research questions, it still means that the researcher is restricted to using either the normalised or original forms and acquiring the original forms of normalised search results can be difficult. The solution adopted in the present edition is to preserve all original spellings in the base data file of the edition and to include normalised forms—together with word-class information—of each word in a separate annotation overlay, linked to the original form by an explicit identifier reference. This allows the use of either original or normalised forms—or even a combination of the two and the word-class data—for queries and the presentation of any results in either original or normalised spelling.

The requirement to keep all annotation separate from ‘the text itself’ is an established one for corpus annotation (see e.g. Leech 2005). In traditional corpora, where all of the textual content can be encoded using Unicode or even ASCII characters and all annotation is analytical and thus ontologically different from ‘the text itself’, this is relatively straightforward. However, descriptive annotation of the textual and paratextual features—as well as special symbols or graphical elements not covered by any character encoding—means that the ontological distinction between annotation and ‘the text itself’ is less clear-cut. Consequently, the editorial framework described in this thesis does not see a ‘plain text’ encoding as a sufficient representation of ‘the text itself’, but sees the crucial demarcation line to lie between the *descriptive annotation* of what is in the document—including the transcription of its textual content—and the *analytical annotation* of what the descriptively annotated features *mean* on a pragmatic or semantic level. The principal means of distinguishing between these two levels of annotation in the present approach is to include the descriptive annotation and the textual transcription in the *base data file* and the various layers of analytical annotation in separate annotation overlays linked to the base data file. The only exception to this is the basic level of text-structural annotation that is included in the base data file and used to organise the textual content into a hierarchical structure extending down to the level of explicitly annotated word-units that are used as the primary *textual coordinate system* for anchoring the analytical annotation layers. Although the inclusion

of the text-structural annotation in the base data file means that the separation of the two annotation layers is not fully realised on the level of file structure, it can easily be replaced by semantically neutral segments or even removed completely from the base data file.<sup>2</sup>

Unlike printed texts, which are inherently closed and stable, the capability of digital resources for cumulative *expansion* has here been argued to be one of their most important benefits. However, most traditional corpora and digital editions do not make use of this capability, but are instead conceived as finished products, much like their printed predecessors. One frequent reason for this is the desire to preserve the integrity and stability of the texts, which is a central requirement of scientific reproducibility. Thus, any mechanism for the continuous expansion of a digital research resource, either in terms of their *breadth*, i.e. the amount of textual data, or their *depth*, i.e. the amount of analytical metadata used to describe the data, will have to include a way of ensuring the integrity and stability of existing data and of documenting the additions. In the present edition this has been accomplished in terms of breadth by storing the descriptive model of each original document as a separate unitary and unchanging base data file, allowing the extent and contents of a corpus to be documented by a list of base data files, and in terms of depth by adding all additional analytical annotation as separate *annotation overlays* whose contents are linked to the data but do not alter it.

These five features of the editorial scheme described in this thesis, together with its diplomatic and graphemic level of annotation, will allow not only the application of a wide variety of corpus-linguistic methods to the data and metadata contained in the data archive, but will also facilitate the cumulative and collaborative production of knowledge.

## 14.2 Cumulative production of knowledge

Perhaps the most important aspect of the editorial framework outlined in this thesis is its emphasis on the *cumulative production of knowledge*, based on scholarly specialisation, division of labour, and recycling of research results to produce new metadata. This concept forms the basis for practically all of the aspects of the edition that could be considered ‘original’ to some degree—although most of them have been suggested by earlier scholars in one form or another. These include the concept of multi-layered annotation, the use of annotation overlays, and the textual coordinate system used to anchor them. The importance of this concept stems largely from its centrality to the development of historical corpora from closed single-purpose products into open and flexible research resources that can be used to find answers to a wide variety of linguistic, textual and even historical research questions.

The production of these kinds of richly annotated and contextualised editions of historical documents and of corpora based on them—whether focused on a specific subject matter, genre, text type, time period or authorial *oeuvre*—requires a variety of different kinds of linguistic, historical, analytical and technical expertise

<sup>2</sup> However, the complete removal of textual structure would also entail the loss of the textual coordinate system and the link between the text and the layers of analytical annotation, and is not recommended.

and a considerable amount of work. For this reason, it is here argued that the best way of bringing about such research resources is to divide the different aspects of the production process—the modelling of historical documents as digital data, analysing these models to create new metadata, and compiling together structured selections of such models to form corpora—between several scholars or teams of scholars, each with their own area of expertise, and over several temporally consecutive projects. This would allow the original edition to form the basis of what Palmer (2004) has called a *thematic research collection*, an open-ended research resource that would “have a potential to grow and change” as scholars “add to and improve the content”, allowing the work on the resource to potentially continue over generations of scholars (351).

This kind of an open-ended resource, however, would require collaboration between scholars preferably from diverse fields—editors, linguists, and historians—as well as co-operation with libraries, museums and other holding institutions that are the custodians of both the original material to be edited as well as different kinds of archives and collections that can be used to contextualise and explain the edited textual objects. This kind of collaboration, although in some large research projects possible within a single project infrastructure, would in most cases involve scholars sharing with each other not only the results of their research in the form of published papers, but also their ‘raw data’ and research process in the form of analytical annotation of the analysed texts. In the field of corpus linguistics, such collaborative thinking has been advocated for example by Leech (2005), who has urged corpus compilers and other linguists annotating corpora to share their linguistic annotations—such as POS-tagging or lexical annotation, which in most cases involve at least some manual work—with the users of a corpus, arguing that not only does this allow others to replicate their research, but also allows the annotation to be used for purposes not imagined by the original annotators and to serve as the basis for further research, resulting in yet new layers of annotation: “In short, an annotated corpus is a sharable resource, an example of the electronic resources increasingly relied on for research and study in the humanities and social sciences.” A similar sentiment was also expressed by Grund in the conclusion of his 2006 article, where he hoped that “in the future, editors and corpus compilers can work together to produce electronic tools, whether in the form of editions or corpora, that can be used for a wide range of linguistic, literary, historical, and paleographic research” (Grund 2006: 122).

However, as Johansson (2004) has observed, many philologists and editors “still automatically object to someone else using ‘their’ text before it has been edited and printed”, seeing the manuscript texts exclusively as material for preparing an edition and objecting to someone else benefiting from the “cumbersome and time-consuming work they have put into their transcriptions and excerpts” (95). Similarly, Wynne (2005) has criticised the reluctance among corpus compilers to make their corpora publicly available, listing some of the reasons for this given by corpus compilers, including:

- 1) to avoid copyright and other rights issues;
- 2) to ensure that the creator has the first, or even exclusive, opportunity to exploit the resource and publish research or further resources based on it;
- 3) to retain the option to sell the rights on a commercial basis;

- 4) because of the danger of uncontrolled commercial exploitation or pirating;
- 5) because it is too much trouble to administer distribution.

In commenting on these reasons, Wynne points out that the first of these is not sound from a legal point of view, as the corpus will or will not constitute a breach of copyright regardless of whether it is publicly distributed or not (although the probability of incurring legal response is of course increased by public distribution). In terms of the second one, he points out that not only do the compilers of a corpus have a head start over other scholars in any case, being able to publish research at the same time as the corpus, but the unavailability of the corpus can hinder the replicability and thus credibility of their research results. As to the third reason, Wynne not only points out that open access and commercial distribution are not necessarily mutually exclusive, but also presents several good arguments for why attempts at the commercial exploitation of language corpora are likely to prove less profitable than open access distribution:

While a commercial deal may please your employer, and bring some financial reward, there are some good arguments for open access. The more widely available the corpus is, the more widely known it is, and the more publicity the creator will receive. A community of researchers who work on the corpus will come into being, creating a higher profile for research based on the resource, including your own. Feedback will be obtained on the usefulness of the resource, and errors can be corrected. Others are more likely to share their resources with you if you share yours. Funders are more likely to give you more funding if you have a good record of ensuring that resources which you have created are properly archived and distributed. The funders generally perceive better value for money in creating resources that are reusable. Failure in this respect could seriously weaken a proposal for further funding. Further project funding may be more lucrative and prestigious than what can be obtained from commercial exploitation of the data.

(Wynne 2005)

For the fourth and fifth reasons, he points out that distribution through an established archive with a rights management policy in fact has a much better chance of controlling access and defending the rights of stakeholders than the corpus compilers themselves, and can also take care of the task of administering the distribution of the material. This reluctance to share the data underlying their research is not, however, limited to corpus linguists. As Cummings (2009) has pointed out, even producers of digital editions and other digital research resources based on the *TEI Guidelines*—in themselves designed to facilitate the sharing of digital texts in a standard format—often pay only lip service to the open sharing of data. While producers of TEI-based resources have often expressed “a publicly declared willingness to share one’s underlying data” (316), there are in fact only a limited number of resources that expose their XML source files in addition to the rendered HTML versions, the excuses for not making the underlying XML available ranging from “licensing concerns to delicate academic egos” (316).

As Cummings (2009) also points out, this means doing a disservice to the whole community, since “as community-driven principles, the TEI Guidelines are only as good as the documents which follow them”, and in order to improve them, we

need examples of actual community practice, both good and bad. This view is also one of the key motivating factors behind the complex nature of this thesis as a specific type of edition of six medieval manuscripts on one level, and a methodological discussion about the principles and practices of producing such an edition on another. The two aspects of this thesis are seen to exist in a symbiotic relationship, the theoretical and methodological discussion supporting and justifying the practical solutions adopted for the edition, and the edition supporting and exemplifying the arguments presented in the discussion. The publication of the full source data for the edition, as well as of the scripts used to generate the three kinds of editorial output (included as appendices B, C and D) from this source data, is intended as an explicit response to the call by Cummings (2009) for editors “to show all of their dirty laundry” (Cummings 2009: 316) for the benefit of the TEI and the wider digital editing community. Regardless of whether the material published here will provide an inspirational or a cautionary example, this solution will hopefully encourage similar openness in other editors.

Although quoted here for his apt and forceful statement of the position, Cummings is not alone in advocating not just an Open Access but also an Open Source approach to the creation of digital research resources. For example Bodard and Garcés (2009) have argued that “a protocol for collections of digital critical editions of texts and/or manuscripts, which aims to allow for collaboration on the widest possible scale, must include the requirement (or at least the very strong recommendation) that texts are not only Open Content (allowing free access to the output itself) but also Open Source—revealing transparently the code behind the output, the research behind the text, the decisions which are part of scholarly publication” (87). According to them, the question of open source data is not merely an economical one, but lies at the heart of the scientific method:

Open source is not so much a business model with exclusively economic implications as a strategy based on the belief that cultural advances are made by building upon the creations and publications of those who came before us. Without full access to the raw code, the documentation and the methodological statement that makes an experiment or a solution reproducible, a given publication is a dead end; it cannot be built upon. (Bodard and Garcés 2009: 86)

Garretson (2008: 76-7) has argued for very similar principles for the development of corpus-linguistic software tools, emphasising the importance of modular design, recycling of existing code and open sharing of code produced in developing the tool, flexible design that allows the addition of new functionality, and detailed documentation of the software.

This emphasis on the collaborative and cumulative development of editions and other digital scholarly resources and on the ethos of openness and sharing that it requires means that in addition to its philological and methodological dimensions, this thesis also has a political dimension, placing itself firmly on the side of Open Access publishing and Open Source development of research resources. In addition to the increased availability and usability of research resources, the distribution of the ‘source code’ of digital editions and other research resources could also have the beneficial side effect of making the technical design and innovation involved in the production of research resources more visible and increasing their



chances of being recognised as a valid source of academic merit:

Like other scholarship in the humanities, research takes place in the production of the resource, and research is advanced as a result of it. Thus, scholarship is embedded in the product and its use. And like research generated in the fields of engineering, computer science, and information science, some of the research contribution lies in the technical design, functionality, and innovation that makes new kinds of research possible. (Palmer 2004: 352)

### 14.3 The *Pottage Dyvers* family

In addition to being a methodological exploration of an editorial approach fulfilling the needs of corpus linguistics, the edition forming the core of this thesis is also intended to make a genuine philological contribution to the study of Middle English culinary recipes by providing a culturally contextualised parallel-text edition of the 15<sup>th</sup>-century family of recipe collections known as *Pottage Dyvers* after the grouping of MSS containing medieval English culinary recipes made in Hieatt (1992: 21). As was observed in the comparison of the *PD* family to other families and individual MSS of medieval English recipes in chapter 9, the *PD* family is not a clearly delimited or unambiguous group, but could also be considered to include other manuscripts than the six traditionally ascribed to it and edited here. In general, the limits between the different Middle English recipe collections as *works* were permeable, sequences of recipes being copied from several collections to make up a new one, making it impossible to classify them exhaustively and unambiguously into families.

There are in fact significant differences between the six manuscript versions edited here, especially between the  $\beta$  group (MSS As and H279 and the other four MSS ( $\gamma$  group)), half of the recipes in the former group not occurring in any of the versions in the latter one. This means that for example Oxford, Bodleian Library MS e. Mus. 52 and London, British Library MS Sloane 1108, described in Hieatt (2008) and compared to the *PD* family in section 9.3, could in fact be included in the family with as much justification as MSS As and H279. The large number of shared recipes makes these two collections the most natural candidates for a future extension of the present edition, and the analysis of their structure in relation to the conclusions of chapter 13 will most likely provide interesting results. Based on the preliminary analysis by Hieatt (2008) of the recipes contained in these collections, they would also seem to contain a significant number of recipes from both the 14<sup>th</sup>-century collections edited by Hieatt and Butler (1985) (MS Sloane 1108) and from the *An Ordinance of Pottage (OP)* family of collections (MS Rawlinson D 1222), indicating that all of these families of collections circulated simultaneously within the same discourse community and were freely combined into new hybrid collections.

In terms of their dialect, all of the six *PD* versions would seem to represent a broadly southern language variant. While there are clear differences in their linguistic profiles, there is also a relatively large body of forms shared by all of the versions, consisting mostly of dialectally neutral forms with general currency, most likely indicative of a relatively advanced state of linguistic standardisation

rather than of the dialectal features of any shared ancestor. As was hypothesised in chapter 6, the language forms of most 15<sup>th</sup>-century texts, including the *PD*, no longer represent distinct individual dialects but rather hybrid accommodations to the different idiolects used in the discourse communities in which they circulate, employing a selection of commonly used forms that are likely to be understandable both to their copyists and to their intended audiences. On the other hand, the fact that there are still clear differences in the dialectal profiles of the six versions despite their relatively late date reveals that standardisation was still far from complete, and most English words still had several morphological and orthographical variants even if they no longer marked consistent dialectal differences. Instead of standardisation in the sense of homogenisation, the indiscriminate mixture of orthographically distinctive parallel forms in the texts would seem to reflect the normalisation of heterogeneous idiolects and the weakening of dialectal identity typical to historical contexts characterised by frequent migration and mixing of geographically diverse populations. In terms of scribal practice this means that instead of translating the orthography of the source or sources into a single homogenous linguistic profile, the scribes seem to be content with employing a variety of forms—often with disparate dialectal origins—for the same word, often even within a single recipe.

Based on the comparison of the aggregate quantitative similarity of dialectal forms between the different versions, calculated in subsection 12.3.1, to the most likely areas of origin determined for each version using the ‘fit’ method in section 12.2, the aggregate similarity between individual versions does not seem to be a very good predictor of the geographical distance between their likely areas of origin. This lack of correlation highlights the variable dialectal significance of the items of the *LALME* questionnaire from one text to another and the ensuing difficulty of quantifying the dialectal similarity of several pairs of linguistic profiles in comparable terms.<sup>3</sup>

While it does not seem possible to distinguish recipe groups with differing origins solely on the basis of linguistic evidence, the analysis of internal variation in the dialectally diagnostic forms within the manuscript versions would in general terms seem to support the conclusions of the structural analysis undertaken in chapter 13, namely that all of the *PD* versions combine recipes originating in different sources.<sup>4</sup> While all of the six surviving copies of the *PD* show signs of having been copied in different parts of the South and the Midlands—with the exception of Douce 55 which shows a high likelihood of having actually been copied in London—both the linguistic evidence of the relative weakness of their dialectal identity and noticeable level of standardisation, and the historical evidence of many European culinary collections originating in royal households (see subsection 8.4.2) make it likely that the textual history of the *Potage Dyvers* family, if not of all the surviving manuscript versions, is closely associated with the capital,

<sup>3</sup> The variation of the dialectal significance of the diagnostic items from one text to the next means that at the least, the kind of similarity metric used in subsection 12.3.1 would need to be supplemented by relative weights assigned to different items and forms, based on both the geographic distribution of the forms and on their frequency in the text. Finding suitable, relatively objective metrics for both of these variables is a complex task and will obviously require further research.

<sup>4</sup> This means that further research combining the methods of structural and dialectal analysis used in chapters 12 and 13 will mostly yield interesting results and will be pursued in a separate research article building on the work undertaken here.

whose metropolitan language use may also provide an explanation for some of the unexpected dialectal mixtures found in many of the versions.

As was shown in chapter 13, the recipes contained in the different *PD* versions and their organisation within the collection varies significantly. However, also the degree of variation itself and thus the resemblance between different pairs of versions varies, some pairs being almost identical with each other and others sharing less than half of their recipes with each other. In terms of their content and internal structure, the six MS versions clearly fall into three pairs which are more or less *parallel versions* of each other: MSS As and H279, MSS Ad and D, and MSS C and H4016. Of these pairs, the latter two contain almost the same selection of recipes but organise them quite differently, while the first pair—which is considerably longer than the other versions—contain a large number of recipes that do not occur in any other version. In addition to the internal relationships of the six surviving versions, an analysis of the distribution of the 371 unique recipes in them indicates that in addition to the 89 ‘core’ recipes that occur in all three of the parallel pairs and would thus seem to originate in a shared ‘archetype’ of the *PD* collection, labeled  $\alpha$  in chapter 13, there would seem to be at least two other major and two minor sources from which recipes originate.

The occurrence of recipes from multiple sources in all of the surviving versions of the *PD* clearly indicates the inadequacy—argued in chapter 3—of the traditional unilineal stemmatic model for representing the medieval copying practices used for these kinds of discourse colony texts, characterised by the combination of multiple sources and frequent reorganisation of the component texts. However, the survival of two almost identically organised copies for each of the three pairs of *PD* versions indicates that in addition to creative reorganisation and combination of sources—resulting in what are essentially new collections—these collections were also being copied with no changes to their content or organisation. This seemingly binary approach to their copying highlights the difference between the roles of the *author* (or even *principal*) and the *animator*, outlined in chapter 2. A copyist fulfilling an *author* (or even a *principal*) role on the level of the entire discourse colony would combine recipes from several sources and organise them into a new version or even an entirely new work, while a copyist restricting himself to the role of an *animator* would merely copy the recipes in the same order as they occur in the exemplar.<sup>5</sup> The apparent prevalence of *authorial* copying in this sense makes it unlikely that the medieval producers and users of recipe collections would have seen these families of collections as unitary *works*, but rather as repositories of knowledge which could be either copied in their original form, either wholly or partially, or mined for material for compiling a new collection for the specific needs of the copyist or his patron.

The fact that recipes seemingly originating in different sources intermingle quite freely in all of the six surviving versions—most likely due to one or more rounds of ‘authorial’ copying—could also help explain the diffuse nature of the dialectal variation observed within individual versions in section 12.4. Since groups of recipes originally derived from the same source no longer occur in discernible groups but are intermingled with recipes from other sources, any dialectal features

<sup>5</sup> This does not mean that this kind of a scribe could not function as an author on the level of individual recipes, linguistically reformulating the contents of the individual recipes.

typical to a specific source are also diffused across the collection and do not occur in spatially distinct sections of the collection. This means that for analysing the dialectal identity of such discourse colonies with diverse sources, the next natural step—which will unfortunately have to be relegated to a separate research article—is to combine these two types of analysis and construct separate linguistic profiles for the recipe groups identified as originating from different sources in order to evaluate the degree to which the dialect of each individual group would seem to be influenced by their individual scribe on the one hand and their textual source on the other.

In light of both the internal relationships of the different *PD* versions edited and analysed here, and their relationships to the other known collections of Middle English recipes, the textual universe of Middle English culinary writing could be more accurately described not as a number of canonical collections occasionally borrowing material from each other, but rather as a vast pool of hundreds—if not thousands—of individual recipes with relatively independent textual histories, with the surviving recipe collections representing more or less established but ultimately ephemeral selections from this pool. While specific collections of recipes clearly did establish themselves sufficiently to occasionally be copied as more or less complete wholes, the occurrence of the same sequences of recipes in several different collections—whether as more or less stable blocks or dispersed among other recipes—indicates that even these ‘established’ collections could be broken apart and merged with material from other collections, reflecting either individual preferences or wider developments in culinary tastes.

From an editorial and philological point of view this means that an individual manuscript recipe collection—much less a critical edition of an entire family of multiple MS collections—is not in fact the most natural unit for editing or studying medieval culinary recipes. The prototypically *rhizomatic* structure of the textual tradition of Middle English culinary writing and the highly unstable nature of the recipe collection as a textual unit mean that both the textual history and the linguistic evolution of medieval culinary writing needs to be studied on the level of the individual recipe. Instead of separate editions of seemingly discrete recipe collections, a more accurate representation of the textual universe of Middle English recipes—as well as other types of texts that frequently occur as parts of *discourse colonies*—would in fact be an interconnected *hypertext corpus* in which the text of each recipe version would not only be embedded in a documentary representation of the specific manuscript collection in which it occurs—representing its *syntagmatic context*—but also linked to all of its parallel versions—representing its *paradigmatic context*. This kind of a corpus—consisting essentially of a series of digital diplomatic editions, linked together on the level of individual recipes—would no longer be exclusively linguistic, but would constitute a versatile multidisciplinary research resource that could be used not only to study the Middle English recipe tradition from a variety of linguistic, textual and historical points of view, but also to serve as a nexus for linking together the different kinds of analytical metadata resulting from these studies in the form of new annotation overlays, potentially bringing about new kinds of *emergent functionality* and opening up avenues for entirely new types of research.

# Bibliography

- A Catalogue of the Harleian Manuscripts in the British Museum with indexes of persons, places, and matters*. 1808. 3 vols. Printed by command of His Majesty King George III. In pursuance of an address of The House of Commons of Great Britain. The British Museum.
- A Fifteenth Century Cookry Boke*. 1962. Comp. by John L. Anderson. illustrator Adrienne Adams. New York: Charles Scribner's Sons.
- Aarseth, Espen. 1997. *Cybertext: Perspectives on Ergodic Literature*. Baltimore, MD: The Johns Hopkins University Press.
- Aarts, Jan and Willem Meijs, eds. 1984. *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research*. Amsterdam: Rodopi.
- Dictionnaire de l'Académie française*. 2005. 9th ed. Paris: A. Fayard. Available at: <http://www.academie-francaise.fr/dictionnaire/index.html>. Accessed 25th July 2012.
- Allen, Rosamund. 1987. "Some Sceptical Observations on the Editing of *The Awntyrs off Arthure*". In: *Manuscripts and Texts: Editorial Problems in Later Middle English Literature*. Ed. by Derek Pearsall. Cambridge: D. S. Brewer, 5–25.
- Alonso Almeida, Francisco. 2001. "Punctuation practice in a late medieval English medical remedybook". In: *Folia Linguistica Historica* 22, 207–37.
- Ambrosoli, Mauro. 1983. "L'Opus Agriculturae di Palladio: volgarizzamenti e identificazione dell'ambiente naturale fra tre e cinquecento". In: *Quaderni storici* XVIII, 227–54.
- 1997. *The wild and the sown: botany and agriculture in Western Europe 1350–1850*. Trans. by Mary Salvatorelli. Cambridge: Cambridge University Press.
- Amos, Mark Addison. 2001. "For Manners Make Man': Bourdieu, de Certeau, and the Common Appropriation of Noble Manners in the *Book of Courtesy*". In: *Medieval Conduct*. Ed. by Kathleen Ashley and Robert L. A. Clark. Minneapolis: University of Minnesota Press, 23–48.
- Anderson, Jean G. et al. 2011. "Data mining and search enhancements using the historical thesaurus of English". In: *SMER '11: Proceedings of the 1st international workshop on Search and mining entity-relationship data*. New York: Association for Computing Machinery, 27–8.
- Arnold, David. 2008. "Digital Artefacts. Possibilities and Purpose". In: *The Virtual Representation of the Past*. Ed. by Mark Greengrass and Lorna Hughes. Farnham: Ashgate, 158–70.
- Association for Documentary Editing - Committee on Electronic Standards. 2002. *Minimum Standards for Electronic Editions*. The Association for Documentary

- Editing. Available at: <http://www.documentaryediting.org/resources/about/standards.html>. Accessed 4th January 2012.
- Atwell, Eric. 2008. "Development of tag sets for part-of-speech tagging". In: *Corpus Linguistics: An International Handbook*. Ed. by Anke Lüdeling and Merja Kytö. Berlin: Walter de Gruyter, 501–27.
- Austin, Thomas, ed. 1888. *Two Fifteenth-Century Cookery Books. Harleian MS. 279 (ab. 1430), & Harl. MS 4016 (ab. 1450), with Extracts from Ashmole MS. 1429, Laud MS. 553, & Douce MS. 55*. Early English Text Society, Original Series 91. London: N. Trübner & Co.
- Bailey, Richard W. 2004. "The need for good texts: the case of Henry Machyn's Day Book". In: *Studies in the history of the English language II: unfolding conversations*. Ed. by Anne Curzan and Kimberly Emmons. Berlin: Mouton de Gruyter, 217–28.
- Baker, Peter S. 1998. "The Reader, the Editor, and the Electronic Critical Edition". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 263–83.
- Bakhtin, Mikhail M. 1986. *Speech Genres and Other Late Essays*. Trans. by Vern W. McGee. Austin, TX: University of Texas Press.
- Barthes, Roland. 1979. "From Work to Text". In: *Textual Strategies: Perspectives in Post-Structuralist Criticism*. Ed. by Josué V. Harari. Ithaca: Cornell University Press, 73–81.
- Barton, David, ed. 1994. *Literacy: an introduction to the ecology of written language*. Oxford: Blackwell.
- Barton, David and Mary Hamilton. 2000. "Literacy practices". In: *Situated Literacies: Reading and Writing in Context*. Ed. by David Barton, Mary Hamilton and Roz Ivanič. London: Routledge, 7–15.
- Baugh, Albert C. and Thomas Cable, eds. 2002. *A History of the English Language*. 5th ed. London: Routledge.
- Bauman, Syd. 2011. "Interchange vs. Interoperability". In: *Proceedings of Balisage: The Markup Conference 2011*. Presented at Balisage: The Markup Conference 2011, Montréal, Canada, August 2 - 5, 2011. Available at: <http://www.balisage.net/Proceedings/vol17/html/Bauman01/BalisageVol17-Bauman01.html>. Accessed 9th May 2012.
- Bazell, Charles Ernest. 1966. "Linguistic typology". In: *Five inaugural lectures*. Ed. by Peter Strevens. Oxford: Oxford University Press, 27–49.
- Beadle, Richard. 2008. "Sir John Fastolf's French Books". In: *Medieval Texts in Context*. Ed. by Graham D. Caie and Denis Renevey. London: Routledge, 96–112.
- Beaugrande, Robert de. 1993. "'Register' in discourse studies: a concept in search of a theory". In: *Register Analysis: Theory and Practice*. Ed. by Mohsen Ghadessy. London: Pinter Publishers, 7–25.
- Beaugrande, Robert de and Wolfgang Ulrich Dressler. 1981. *Introduction to text linguistics*. London: Longman.
- Bennett, Michael. 2009. "France in England: Anglo-French Culture in the Reign of Edward III". In: *Language and Culture in Medieval Britain: the French of England c. 1100-1500*. Ed. by Jocelyn Wogan-Browne et al. York: York Medieval Press, 320–33.
- Benskin, Michael. 1991. "The 'fit'-technique explained". In: *Regionalism in late medieval manuscripts and texts: essays celebrating the publication of A linguistic*

- atlas of late mediaeval English. Ed. by Felicity Riddy. Cambridge: D. S. Brewer, 9–26.
- Benskin, Michael and Margaret Laing. 1981. "Translations and Mischsprachen in Middle English Manuscripts". In: *So meny people longages and tonges: Philological Essays in Scots and Mediaeval English Presented to Angus McIntosh*. Ed. by Michael Benskin and M. L. Samuels. Edinburgh: Middle English Dialect Project, 55–106.
- Benskin, Michael, Margaret Laing et al. 2013. *An Electronic Version of A Linguistic Atlas of Late Mediaeval English*. The University of Edinburgh. Available at: <http://www.lel.ed.ac.uk/ihd/elalme/elalme.html>. Accessed 17th November 2013.
- Benson, Larry, ed. 1988. *The Riverside Chaucer*. 3rd ed. Oxford: Oxford University Press.
- Bernard, Edward. 1697. *Catalogi librorum manuscriptorum Angliae et Hiberniae in unum collecti cum indice alphabetico*. Vol. 2. Oxoniae: E Theatro Sheldoniano. Facsimile copy provided by *Early English Books Online (EEBO)*.
- Berrie, Phill et al. 2006. "Authenticating electronic editions". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 269–76.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Biber, Douglas, Susan Conrad and Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, Douglas and Edward Finegan, eds. 1994. *Sociolinguistic Perspectives on Register*. Oxford: Oxford University Press.
- BL MSS Catalogue. *British Library Manuscripts Catalogue*. The British Library. Available at: <http://www.bl.uk/catalogues/manuscripts/INDEX.asp>. Accessed 31st January 2012.
- Black, William Henry. 1845. *A Descriptive, Analytical, and Critical Catalogue of the Manuscripts Bequeathed unto the University of Oxford by Elias Ashmole, Esq., M.D., F.R.S., Windsor Herald, also of some additional Mss. contributed by Kingsley, Lhuyd, Borlase, and others*. Oxford: Oxford University Press.
- Blake, Norman. 1992a. "Geoffrey Chaucer: Textual Transmission and Editing". In: *Crux and Controversy in Middle English Textual Criticism*. Ed. by Alistair J. Minnis and Charlotte Brewer. Cambridge: D. S. Brewer, 19–38.
- 1992b. "Translation and the history of English". In: *History of Englishes. New Methods and Interpretations in Historical Linguistics*. Ed. by Matti Rissanen et al. Berlin: Mouton de Gruyter, 3–24.
- 1998. "Reflections of the Editing of Middle English Texts". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 61–77.
- 2000. "A New Approach to the Witnesses and Text of the *Canterbury Tales*". In: *New Directions in Later Medieval Manuscript Studies*. Ed. by Derek Pearsall. Rochester, NY: York Medieval Press, 29–40.

- Bodard, Gabriel and Juan Garcés. 2009. "Open Source Critical Editions: A Rationale". In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 83–98.
- Boffey, Julia and John J. Thompson. 1989. "Anthologies and miscellanies: production and choice of texts". In: *Book Production and Publishing in Britain 1375–1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 279–315.
- Bolter, Jay David. 1991. *Writing Space: The Computer, Hypertext, and the History of Writing*. Hillsdale, N.J.: Lawrence Earlbaum Associates.
- Bos, Bert et al., eds. 2011. *Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification*. W3C Recommendation. The World Wide Web Consortium. Available at: <http://www.w3.org/TR/2011/REC-CSS2-20110607/>. Accessed 27th February 2012.
- Bowers, Fredson. 1959. *Textual & Literary Criticism*. Cambridge: Cambridge University Press.
- 1972. "Multiple Authority: New Problems and Concepts of Copy-Text". In: *The Library* 27, 81–115.
- 1989. "Regularization and normalization in modern critical texts". In: *Studies in Bibliography* 42, 79–102.
- Brand, Paul. 2000. "The languages of the law in later medieval England". In: *Multilingualism in Later Medieval Britain*. Ed. by David Trotter. Cambridge: D. S. Brewer, 63–76.
- Braswell, Laurel. 1984. "Utilitarian and Scientific Prose". In: *Middle English Prose: a critical guide to major authors and genres*. Ed. by A. S. G. Edwards. New Brunswick, NJ: Rutgers University Press, 337–87.
- Braudel, Fernand. 1981. *Civilization and capitalism: 15th–18th century*. Vol. 1: *The structures of everyday life: the limits of the possible*. Trans. by Siân Reynolds. London: Fontana Press.
- Brears, Peter. 2012. *Cooking & dining in Medieval England*. Totnes: Prospect Books.
- Bree, Linda and James McLaverty. 2009. "The Cambridge Edition of the Works of Jonathan Swift and the Future of the Scholarly Edition". In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 127–36.
- Brewer, Charlotte. 1992. "George Kane's Processes of Revision". In: *Crux and Controversy in Middle English Textual Criticism*. Ed. by Alistair J. Minnis and Charlotte Brewer. Cambridge: D. S. Brewer, 71–96.
- Briquet, Charles-Moïse. 1923. *Les Filigranes: Dictionnaire historique des marques du papier del leur apparition vers 1282 jusqu'en 1600*. 2nd ed. 4 vols. Leipzig: Verlag Von Karl W. Heirsmann.
- Bristow, Whiston. 1764. *Musæum Thoresbyanum. A catalogue Of the genuine and valuable collection Of that well known Antiquarian the late Ralph Thoresby, Gent. F. R. S. Author of Ducatus Leodiensis. Consisting of Roman, British, Runic, Saxon, and English Coins and Medals in Gold, Silver, &c. Manuscripts, Curiosities, Autographs, antient Deeds, original Letters and Signs Manual of British and foreign Kings and Queens, Cromwell the Protector and his Son Richard, Principal Nobility and eminent Persons, for Two Centuries past. All which will be sold by auction, By Whiston Bristow, Sworn Broker, At the Exhibition Room, Spring Gardens, Charing Cross, On Monday March 5th, and the Two following Days, begining punctually at 12 O'Clock. To be viewed on Thursday March 1st, and the following Days,*



- Sunday excepted*. London. Facsimile copy provided by *Eighteenth Century Collections Online (ECCO)*.
- Britnell, Richard. 1997. "Pragmatic Literacy in Latin Christendom". In: *Pragmatic Literacy East and West, 1200–1300*. Ed. by Richard Britnell. Woodbridge: Boydell Press, 3–24.
- 2009. "Uses of French Language in Medieval English Towns". In: *Language and Culture in Medieval Britain: the French of England c. 1100–1500*. Ed. by Jocelyn Wogan-Browne et al. York: York Medieval Press, 81–9.
- Brockbank, Philip. 1991. "Towards a mobile text. Essays in Honour of James T. Boulton". In: *The Theory and Practice of Text-Editing*. Ed. by Ian Small and Marcus Walsh. Cambridge: Cambridge University Press, 90–106.
- Brown, Carleton and Rossell Hope Robbins. 1943. *The Index of Middle English Verse*. New York: Columbia University Press.
- Burnard, Lou. 2005. "Metadata for Corpus Work". In: *Developing Linguistic Corpora: a Guide to Good Practice*. Ed. by Martin Wynne. Oxford: Oxbow Books, 30–46. Available at: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter3.htm>. Accessed 29th January 2014.
- Burnley, John David. 2001. "French and Frenches in fourteenth-century London". In: *Language Contact in the History of English*. Ed. by Dieter Kastovsky and Arthur Mettinger. Frankfurt am Main: Peter Lang, 17–34.
- Burrow, John A. 1988. *The ages of man: a study in medieval writing and thought*. Oxford: Clarendon Press.
- Butcher, Andrew. 2004. "The functions of script in the speech community of a late medieval town, c. 1300–1550". In: *The Uses of Script and Print, 1300–1700*. Ed. by Julia Crick and Alexandra Walsham. Cambridge: Cambridge University Press, 157–70.
- Buzzetti, Dino. 2004. "Diacritical ambiguity and markup. Proceedings of a conference at Bologna, 22–23 September 2002". In: *Augmenting Comprehension: Digital Tools and the History of Ideas*. Ed. by Dino Buzzetti, Giuliano Pancaldi and Harold Short. London: Office for Humanities Communication, 175–88.
- 2009. "Digital Editions and Text Processing". In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 45–61.
- Buzzetti, Dino and Jerome McGann. 2006. "Critical editing in a digital horizon". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 53–73.
- Bühler, Curt F. 1960. *The 15<sup>th</sup> century book: the scribes, the printers, the decorators*. Philadelphia: University of Pennsylvania Press.
- Bynum, Caroline Walker. 1987. *Holy Feast and Holy Fast: The Religious Significance of Food to Medieval Women*. Berkeley: University of California Press.
- Bøe, Hilde, Jon Gunnar Jørgensen and Stine Brenna Taugbøl. 2004. "Philology Meets Text Encoding in the New Scholarly Edition of Henrik Ibsen's Writings". In: *Literary and Linguistic Computing* 19, 55–71.
- Cabrera-Abreu, Mercedes. 2002. "The Formulation of Promise in Medieval English Medical Recipes: A Relevance-Theoretic Approach". In: *Neophilologus* 86, 137–54.

- Caie, Graham D. 2000. "Hypertext and multiplicity: the medieval example". In: *The Renaissance Text: Theory, editing, textuality*. Ed. by Andrew Murphy. Manchester: Manchester University Press, 30–43.
- 2003. "New Corn from Old Fields': the auctor and compiler in fourteenth-century English literature". In: *Revista Canaria de Estudios Ingleses* 47, 59–71. A preprint version available online at <http://eprints.gla.ac.uk/2058/>.
  - 2008. "The manuscript experience: what medieval vernacular manuscripts tell us about authors and texts". In: *Medieval Texts in Context*. Ed. by Graham D. Caie and Denis Renevey. London: Routledge, 10–27.
- Campbell, Lyle, ed. 2004. *Historical Linguistics: An Introduction*. 2nd ed. Edinburgh: Edinburgh University Press.
- Carlquist, Jonas. 2004. "Medieval Manuscripts, Hypertext and Reading. Visions of Digital Editions". In: *Literary and Linguistic Computing* 19, 105–18.
- Carroll, Ruth. 1999. "The Middle-English Recipe as a Text Type". In: *Neuphilologische Mitteilungen* 100, 27–42.
- 2003. "Recipes for laces: An example of a Middle English discourse colony". In: *Discourse Perspectives on English: Medieval to Modern*. Ed. by Risto Hiltunen and Janne Skaffari. Amsterdam: John Benjamins Publishing, 137–65.
  - 2004. "Middle English Recipes: vernacularisation of a text-type". In: *Medical and Scientific Writing in Late Medieval English*. Ed. by Irma Taavitsainen and Päivi Pahta. Cambridge: Cambridge University Press, 174–91.
  - 2006. "Assessing palaeographic evidence for discourse structuring in Middle English recipes". In: *Boletín Millares Carlo* 24–25, 305–25.
  - 2009. "Vague language in the medieval recipes of the *Forme of Cury*". In: *Instructional Writing in English: Studies in honour of Risto Hiltunen*. Ed. by Matti Peikola, Janne Skaffari and Sanna-Kaisa Tanskanen. Amsterdam: John Benjamins, 55–82.
- Cayless, Hugh A. and Adam Soroka. 2010. "On Implementing string-range() for TEI". In: *Proceedings of Balisage: The Markup Conference 2010*. Presented at Balisage: The Markup Conference 2010, Montréal, Canada, August 3 - 6, 2010. Available at: <http://www.balisage.net/Proceedings/vol15/html/Cayless01/BalisageVol15-Cayless01.html>. Accessed 9th May 2012.
- Cerquiglini, Bernard. 1989. *Éloge de la variante: Histoire critique de la philologie*. Paris: Seuil.
- 1999. *In Praise of the Variant*. Trans. by Betsy Wing. Baltimore: Johns Hopkins University Press.
- Červenka, Miroslav. 1995. "Textual Criticism and Semiotics". In: *Contemporary German Editorial Theory*. Ed. by Hans Walter Gabler, George Bornstein and Gillian Borland Pierce. Ann Arbor: University of Michigan Press, 59–77.
- Chartier, Roger. 1995. *Forms and Meanings: Texts, Performances, and Audiences from Codex to Computer*. Philadelphia: University of Pennsylvania Press.
- Chaytor, Henry J. 1950. *From Script to Print: An Introduction to Medieval Vernacular Literature*. London: Sidgwick Jackson.
- Cherry, Martin. 2004. *Courtenay, Thomas, thirteenth earl of Devon (1414–1458)*. In: *Oxford Dictionary of National Biography*. Ed. by Lawrence Goldman. Oxford: Oxford University Press.
- Christianson, Paul. 1990. *A Directory of London Stationers and Book Artisans 1300–1500*. New York: The Bibliographical Society of America.

- Clanchy, M. T. 1993. *From Memory to Written Record: England 1066-1307*. 2nd ed. Oxford: Blackwell Publishers Ltd.
- Claridge, Claudia. 2008. "Historical corpora". In: *Corpus Linguistics: An International Handbook*. Ed. by Anke Lüdeling and Merja Kytö. Berlin: Walter de Gruyter, 242–59.
- Clark, Cecily. 1992. "The myth of 'the Anglo-Norman Scribe'". In: *History of Englishes. New Methods and Interpretations in Historical Linguistics*. Ed. by Matti Rissanen et al. Berlin: Mouton de Gruyter, 117–29.
- Clark, James and Makoto Muraka, eds. 2001. *RELAX NG Specification*. Committee Specification. The Organization for the Advancement of Structured Information Standards. Available at: <http://relaxng.org/spec-20011203.html>. Accessed 23rd January 2014.
- Cohen, Philip and David H. Jackson. 1991. "Notes on Emerging Paradigms in Editorial Theory". In: *Devils and Angels: Textual Editing and Literary Theory*. Ed. by Philip Cohen. Charlottesville: University Press of Virginia, 103–23.
- Cokayne, George E. 1926. *The complete peerage of England, Scotland, Ireland, Great Britain, and the United Kingdom, extant, extinct, or dormant: VI. Gordon to Hustpierpoint*. Ed. by H. A. Doubleday, Duncan Warrand and Lord Howard de Walden. London: St. Catherine Press.
- 1932. *The complete peerage of England, Scotland, Ireland, Great Britain, and the United Kingdom, extant, extinct, or dormant: VIII. Lindley to Moate*. Ed. by H. A. Doubleday and Lord Howard de Walden. London: St. Catherine Press.
- Coleman, Janet. 1981. *English Literature in History 1350-1400: Medieval Readers and Writers*. London: Hutchinson.
- Cosin MSS. 2008. "Collection Level Description: Cosin Manuscripts". In: *Durham University Library ASC Collections information*. Available at: [http://www.dur.ac.uk/library/asc/collection\\_information/cldload/?collno=28](http://www.dur.ac.uk/library/asc/collection_information/cldload/?collno=28). Accessed 24th September 2012.
- Connolly, Margaret. 1992. "The Deth of the Kynge of Scotis: A new edition". In: *The Scottish Historical Review* 71, 46–69.
- Cooper, Helen. 1998. "Averting Chaucer's Prophecies: Miswriting, Mismetering, and Misunderstanding". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 79–93.
- Crane, Greg. 2006. "Document management and file naming". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 277–90.
- Crespo Garcia, Begoña. 2000. "Historical Background of multilingualism and its impact on English". In: *Multilingualism in Later Medieval Britain*. Ed. by David Trotter. Cambridge: D. S. Brewer, 25–36.
- Crombie, Alistair Cameron. 1994. *Styles of scientific thinking in the European tradition: the history of argument and explanation especially in the mathematical and biomedical sciences and arts*. London: Duckworth.
- Culy, Christopher. 1996. "Null objects in English recipes". In: *Language Variation and Change* 8, 141–76.
- Cummings, James. 2009. "Converting Saint Paul: A new TEI P5 edition of *The Conversion of Saint Paul* using stand-off methodology". In: *Literary and Linguistic Computing* 24, 307–17.

- Curzan, Anne and Kimberly Emmons, eds. 2004. *Studies in the history of the English language II: unfolding conversations*. Topics in English Linguistics 45. Berlin: Mouton de Gruyter.
- Curzan, Anne and Chris C. Palmer. 2006. "The importance of historical corpora, reliability, and reading". In: *Corpus-based Studies of Diachronic English*. Ed. by Roberta Facchinetti and Matti Rissanen. Bern: Peter Lang, 17–34.
- Dahlström, Erik et al., eds. 2011. *Scalable Vector Graphics (SVG) 1.1*. 2nd ed. W3C Recommendation. The World Wide Web Consortium. Available at: <http://www.w3.org/TR/2011/REC-SVG11-20110816/>. Accessed 18th January 2012.
- Dahlström, Mats. 2009. "The Compleat Edition". In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 27–44.
- Dalby, Andrew. 2000. *Dangerous tastes: the story of spices*. London: The British Museum Press.
- Damerau, Fred J. 1964. "A Technique for Computer Detection and Correction of Spelling Errors". In: *Communications of the ACM* 7, 171–6.
- Davis, Norman, ed. 1958. *Paston Letters*. Oxford: Clarendon Press.
- De Hamel, Christopher. 1992. *Medieval Craftsmen: Scribes and Illuminators*. London: British Museum Press.
- De Smedt, Marcel and Edward Vanhoutte, eds. 2000. *Stijn Streuvels, De Teleurgang van den Waterhoek. Elektronisch-kritische editie/electronic-critical edition*. CD-ROM. Amsterdam.
- Deegan, Marilyn. 2006. "Collection and preservation of an electronic edition". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keefe and John Unsworth. New York: The Modern Language Association of America, 358–70.
- Deegan, Marilyn and Simon Tanner. 2004. "Conversion of Primary Sources". In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, Ray Siemens and John Unsworth. Malden, MA: Blackwell Publishing, 488–504.
- Dekhtyar, Alex and Ionut E. Iacob. 2003. "A framework for management of concurrent XML markup. ER 2003 Workshops ECOMO, IWCMQ, AOIS, and XSDM, Chicago, IL, USA, October 13, 2003. Proceedings". In: *Conceptual Modeling for Novel Application Domains*. Ed. by Manfred A. Jeusfeld and Óscar Pastor. Springer, 185–208.
- Del Lungo Camiciotti, Gabriella. 2011. "Textuality in late mediaeval England: Two case studies". In: *Communicating Early English Manuscripts*. Ed. by Päivi Pahta and Andreas H. Jucker. Cambridge: Cambridge University Press, 25–37.
- Deleuze, Gilles and Félix Guattari. 1987. *Capitalism and Schizophrenia, vol. 2: A Thousand Plateaus*. Trans. by Brian Massumi. Minneapolis: University of Minnesota Press.
- Denholm-Young, N. 1954. *Handwriting in England and Wales*. Cardiff: University of Wales Press.
- DeRose, Steven. 2004. "Markup Overlap: A Review and a Horse". In: *Proceedings of Extreme Markup Languages Conference, Montreal, Québec, 2004*. Available at: <http://conferences.idealliance.org/extreme/html/2004/DeRose01/EML2004DeRose01.html>. Accessed 9th May 2012.
- DeRose, Steven, Eve Maler and Ron Daniel Jr., eds. 2002. *XPointer xpointer() Scheme*. W3C Working Draft. The World Wide Web Consortium. Available at:

- <http://www.w3.org/TR/2002/WD-xptr-xpointer-20021219/>. Accessed 5th January 2012.
- Di Iorio, Angelo, Silvio Peroni and Fabio Vitali. 2009. "Towards markup support for full GODDAGs and beyond: the EARMARK approach". In: *Proceedings of Balisage: The Markup Conference 2009*. Presented at Balisage: The Markup Conference 2009, Montréal, Canada, August 11 - 14, 2009. Available at: <http://www.balisage.net/Proceedings/vol3/html/Peroni01/BalisageVol3-Peroni01.html>. Accessed 9th May 2012.
- Diehl, Peter. 2004. "An Inquisitor in Manuscript and in Print: The *Tractatus super materia hereticorum* of Zanchino Ugolini". In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, 58–77.
- Dipper, Stefanie. 2008. "Theory-driven and corpus-driven computational linguistics, and the use of corpora". In: *Corpus Linguistics: An International Handbook*. Ed. by Anke Lüdeling and Merja Kytö. Berlin: Walter de Gruyter, 68–96.
- Dollinger, Stefan. 2004. "'Philological computing' vs. 'philological outsourcing' and the compilation of historical corpora: a Late Modern English test case". In: *Vienna English Working Papers (VIEWS)* 13. Ed. by Christiane Dalton-Puffer et al., 3–23.
- Douglas, Mary. 1979. *Purity and danger: an analysis of the concepts of pollution and taboo*. London: Routledge & Kegan Paul.
- Doyle, A. I. 1992. "Cosin MS. V.III.11: Tractatus VI (Miscellany)". In: *Draft Catalogue of Medieval Manuscripts in the University Library*. Available at: <https://www.dur.ac.uk/library/asc/theme/medmss/apviii11/>. Accessed 30th July 2012.
- 2000. "Recent Directions in Medieval Manuscript Study". In: *New Directions in Later Medieval Manuscript Studies*. Ed. by Derek Pearsall. Rochester, NY: York Medieval Press, 1–14.
- 2004. "The Cosin Manuscripts and George Davenport". In: *The Book Collector* 53, 32–45.
- Driscoll, M. J. 2006. "Levels of transcription". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keefe and John Unsworth. New York: The Modern Language Association of America, 254–61.
- Drucker, Johanna. 2014. "Pixel Dust: Illusions of Innovation in Scholarly Publishing". In: *Los Angeles Review of Books*. Available at: <http://lareviewofbooks.org/essay/pixel-dust-illusions-innovation-scholarly-publishing>. Accessed 8th February 2014.
- Duggan, Hoyt N. 1996. "Libertine scribes and maidenly editors: meditations on textual criticism and metrics". In: *English Historical Metrics*. Ed. by C. B. McCully and J. J. Anderson. Cambridge: Cambridge University Press, 219–37.
- Durusau, Patrick. 2006. "Why and how to document you markup choices". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keefe and John Unsworth. New York: The Modern Language Association of America, 299–309.
- Durusau, Patrick and Matthew Brook O'Donnel. 2004. "Tabling the Overlap Discussion". In: *Proceedings of Extreme Markup Languages Conference, Montreal, Québec, 2004*. Available at: <http://conferences.idealliance.org/extreme/html/2004/Durusau01/EML2004Durusau01.html>. Accessed 9th May 2012.

- Dyer, Christopher. 1989. *Standards of living in the later Middle Ages: Social change in England c. 1200–1520*. Cambridge: Cambridge University Press.
- Eamon, William, ed. 1982. *Studies on medieval Fachliteratur. proceedings of the Special Session on Medieval Fachliteratur of the sixteenth International Congress on Medieval Studies, Kalamazoo, Michigan (U.S.A.), May 10, 1981*. Scripta: mediaeval and Renaissance texts and studies 6. Brussel: UFSAL.
- Early English Text Society, ed. 2011. *Guidelines for Editors*. Early English Text Society. Available at: <http://users.ox.ac.uk/~eets/Guidelines%20for%20Editors%2011.pdf>. Accessed 4th 18 2013.
- Eaves, Morris. 2006. "Multimedia body plans: a self-assessment". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 210–23.
- Echard, Siân and Stephen Partridge. 2004a. "Introduction - Varieties of Editing: History, Theory, and Technology". In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, xi–xxi.
- eds. 2004b. *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Toronto: University of Toronto Press.
- Edwards, A. S. G. 1987. "Observations on the History of Middle English Editing". In: *Manuscripts and Texts: Editorial Problems in Later Middle English Literature*. Ed. by Derek Pearsall. Cambridge: D. S. Brewer, 34–48.
- 1998a. "Editing and the Teaching of Alliterative Verse". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 95–106.
- 1998b. "Manuscript and Text". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 159–68.
- 2000. "Representing the Middle English Manuscript". In: *New Directions in Later Medieval Manuscript Studies*. Ed. by Derek Pearsall. Rochester, NY: York Medieval Press, 65–79.
- Edwards, A. S. G. and Julia Boffey. 2005. *A New Index of Middle English Verse*. London: British Library.
- Edwards, A. S. G. and Douglas Moffat. 1998. "Annotation". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 217–36.
- Edwards, A. S. G. and Derek Pearsall. 1989. "The manuscripts of the major English poetic texts". In: *Book Production and Publishing in Britain 1375–1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 257–78.
- Edwards, Anthony William Fairbank. 1989. "Venn diagrams for many sets". In: *New Scientist* 121, 51–6.
- Eggert, Paul. 1991. "Textual Product or Textual Process: Procedures and Assumptions of Critical Editing". In: *Devils and Angels: Textual Editing and Literary Theory*. Ed. by Philip Cohen. Charlottesville: University Press of Virginia, 57–77.
- 1994. "Document and Text: The 'Life' of the Literary Work and the Capacities of Editing". In: *TEXT: Transactions of the Society for Textual Scholarship* 7, 1–24.

- 2004. "The Way of All Text: The Materialist Shakespeare". In: *Voice, Text, Hypertext: Emerging Practices in Textual Studies*. Ed. by Raimonda Modiano, Leroy F. Searle and Peter Shillingsburg. Seattle: University of Washington Press, 162–76.
- 2009. "The Book, the E-text and the 'Work-site'". In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 63–82.
- Eggins, Suzanne, ed. 1994. *An Introduction to Systemic Functional Linguistics*. London: Pinter Publishers.
- Eggins, Suzanne and J. R. Martin. 1997. "Genres and Registers of Discourse". In: *Discourse as Structure and Process*. Ed. by Teun Van Dijk. London: Sage Publications, 230–56.
- Eisenstein, Elizabeth. 1979. *The printing press as an agent of change: communications and cultural transformations in early modern Europe*. Cambridge: Cambridge University Press.
- Elo, Tuulikki, Heikki Laaksonen and Eero Valjakka, eds. 2002. *Apicius: Roomalainen keittokirja - De Re Coquinaria*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Embree, Dan and Elizabeth Urquhart. 1987. "The Simonie: the Case For a Parallel-text Edition". In: *Manuscripts and Texts: Editorial Problems in Later Middle English Literature*. Ed. by Derek Pearsall. Cambridge: D. S. Brewer, 49–59.
- Emden, Alfred B. 1957–59. *A biographical register of the University of Oxford to A.D. 1500*. 3 vols. Oxford: Clarendon Press.
- Expert Advisory Group on Language Engineering Standards (EAGLES). 2000. *Corpus Encoding Standard*. 1.5. Expert Advisory Group on Language Engineering Standards (EAGLES). Available at: <http://www.cs.vassar.edu/CES/>. Accessed 15th May 2014.
- Faccioli, Emilio. 1966. *Arte della cucina, Libri di ricette: Testi sopra lo scalco, il trinciante e i vini dal XIV al XIX secolo*. Milan: Edizioni Il Polifilo.
- Fahy, Conor. 2004. "Old and New in Italian Textual Criticism". In: *Voice, Text, Hypertext: Emerging Practices in Textual Studies*. Ed. by Raimonda Modiano, Leroy F. Searle and Peter Shillingsburg. Seattle: University of Washington Press, 401–11.
- Fairclough, Norman. 1992. *Discourse and Social Change*. Cambridge: Polity Press.
- Fawcett, Richard. 1999. *Stirling Castle*. Edinburgh: Historic Scotland.
- Fellows, Jennifer. 1998. "Author, Author, Author . . . : An Apology for Parallel Texts". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 15–24.
- Fenton, Eileen Gifford and Hoyt N. Duggan. 2006. "Effective methods of producing machine-readable text from manuscript and print sources". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keefe and John Unsworth. New York: The Modern Language Association of America, 241–53.
- Fischer, Olga. 1992. "Syntax". In: *The Cambridge History of the English Language: Volume II, 1066–1476*. Ed. by Norman Blake. Cambridge: Cambridge University Press, 207–408.
- Fisher, John H. 1977. "Chancery and the Emergence of Standard Written English in the Fifteenth Century". In: *Speculum* 52, 870–99.

- Flanders, Julia. 2006. "The Women Writers Project: a digital anthology". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 138–49.
- 2009. "Data and Wisdom: Electronic Editing and the Quantification of Knowledge". In: *Literary and Linguistic Computing* 24, 53–62.
- Flandrin, Jean Louis. 1992. "Structure des menus français et anglais aux XI<sup>e</sup> et XV<sup>e</sup> siècles". In: *Du Manuscrit à la table: Essais sur la cuisine au Moyen Âge et répertoire des manuscrits médiévaux contenant des recettes culinaires*. Ed. by Carole Lambert. Montreal: Les Presses de l'Université de Montreal, 173–92.
- Fleischman, Suzanne. 1986. "Philology, Linguistics, and the Discourse of the Medieval Text". In: *Speculum* 61, 19–37.
- 2000. "Methodologies and Ideologies in Historical Linguistics". In: *Textual Parameters in Older Languages*. Ed. by Susan Herring, Pieter van Reenen and Lene Schøsler. Amsterdam: Benjamins, 33–58.
- Fletcher, Alan J. 1994. "'Benedictus qui venit in nomine domini': A thirteenth-century sermon for Advent and macaronic style in England". In: *Mediaeval Studies* 56, 217–45.
- Foster, Joseph, ed. 1891. *Alumni Oxonienses: The members of the University of Oxford, 1500-1714: their parentage, birthplace, and year of birth, with a record of their degrees*. 4 vols. Oxford: James Parker & Co. Available at: <http://archive.org/details/alumnioxoniensi01fostgoog>. Accessed 6th August 2012.
- Foucault, Michel. 1979. "What Is an Author?". In: *Textual Strategies: Perspectives in Post-Structuralist Criticism*. Ed. by Josué V. Harari. Ithaca: Cornell University Press, 141–60.
- Foulet, Alfred and Mary Blakely Speer. 1979. *On editing Old French texts*. Lawrence: Regents Press of Kansas.
- Fraistat, Neil and Steven Jones. 2006. "The poem and the network: editing poetry electronically". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 105–21.
- Francis, W. Nelson. 1980. "A Tagged Corpus: Problems and Prospects". In: *Studies in English Linguistics: for Randolph Quirk*. Ed. by Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. London: Longman, 192–209.
- Francis, W. Nelson and Henry Kučera. 1964. *Manual of Information to accompany A Standard Corpus of Present-Day Edited American English, for use with Digital Computers*. Revised 1971. Revised and amplified 1979. Rhode Island: Department of Linguistics, Brown University. Available at: <http://icame.uib.no/brown/bcm.html>. Accessed 25th November 2013.
- 1979. *Manual of Information to Accompany a Standard Sample of Present-day Edited American English, for Use with Digital Computers*. Original edition 1964, revised and augmented 1979. Providence: Brown University.
- Fрати, Ludovico. 1899. *Libro di cucina del secolo xiv*. Livorno: Raffaello Giusti. Available at: <http://archive.org/details/librodicucinade00fratgoog>. Accessed 8th October 2013.
- Freeman, Michael. 1977. "Sung". In: *Food in Chinese Culture: anthropological and historical perspectives*. Ed. by K. C. Chang. New Haven: Yale University Press, 141–76.



- Froehlich, Heather. 2013. "A Text Analysis By Any Other Name". In: *hfroehlich.wordpress.com*. Blog entry. Available at: <http://hfroehlich.wordpress.com/2013/01/14/by-any-other-name/>. Accessed 24th January 2014.
- Fryde, E. B. et al. 1996. *Handbook of British Chronology*. Third revised ed. Cambridge: Cambridge University Press.
- Furnivall, Frederick James, ed. 1868. *Early English Meals and Manners: John Russell's Boke of Nurture, Wynkyn de Worde's Boke of Keruynge, The Boke of Curtesye, R. Weste's Booke of Demeanor, Seager's Schoole of Vertue, The Babees Book, Aristotle's A B C, Urbanitatis, Stans Puer ad Mensam, The Lytylle Childrenes Lytil Boke, for to serve a Lord, Old Symon, The Birched School-Boy, &c. &c. with some forewords on Education in Early England*. Early English Text Society, Original Series 32. London: H. Milford, Oxford University Press.
- ed. 1869. *Queene Elizabethes Achademy, A Booke of Precedence, &c. with essays on Italian and German books on courtesye by W. M. Rossetti and E. Oswald*. Early English Text Society, Extra Series 8. London: N. Trübner.
- Futernick, Robert. 1983. "Leaf Casting on the Suction Table". In: *Journal of the American Institute for Conservation* 22, 82–91.
- Gabler, Hans Walter. 2006. "Moving a print-based editorial project into electronic form". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 339–45.
- 2007. "The Primacy of the Document in Editing". In: *Ecdotica* 4. Ed. by Loredana Chines, 197–207. Available at: [http://academia.edu/166634/\\_The\\_Primacy\\_of\\_the\\_Document\\_in\\_Editing.\\_](http://academia.edu/166634/_The_Primacy_of_the_Document_in_Editing._). Accessed 11th December 2012.
- Gabler, Hans Walter, George Bornstein and Gillian Borland Pierce, eds. 1995. *Contemporary German Editorial Theory*. Ann Arbor: University of Michigan Press.
- Gaertner, Otto. 1904. "John Shirley sein Leben und Wirken". PhD thesis. Halle, a.d.S.: Vereinigte Friedrichs-Universität Halle-Wittenberg.
- Gants, David. 2006. "Drama case study: *The Cambridge Edition of the Works of Ben Jonson*". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 122–37.
- Garretson, Gregory. 2008. "Desiderata for Linguistic Software Design". In: *International Journal of English Studies* 8, 67–94.
- Gaskell, Philip. 1978. *From Writer to Reader: Studies in Editorial Method*. Oxford: Clarendon Press.
- Genette, Gérard. 1987. *Seuils*. Paris: Seuil.
- 1997. *Paratexts: Thresholds of Interpretation*. Trans. by Jane E. Lewin. With a forew. by Richard Macksey. Cambridge: Cambridge University Press.
- Gieben, Servus. 1967. "Robert Grosseteste and Medieval Courtesy-Books". In: *Vivarium* 5, 47–74.
- Gillam, S. G., ed. 1984. *The Douce Legacy: An exhibition to commemorate the 150th anniversary of the bequest of Francis Douce (1757-1834)*. Oxford: Bodleian Library.
- Gillespie, Vincent. 1989. "Vernacular books of religion". In: *Book Production and Publishing in Britain 1375-1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 317–44.

- Glare, P. G. W., ed. 2012. *Oxford Latin Dictionary*. 2nd ed. Oxford: Oxford University Press.
- Gleßgen, Martin-Dietrich and Franz Lebensanft. 1997. "Von alter und neuer Philologie: Neuer Streit über Prinzipien und Praxis der Textkritik". In: *Alte und neue Philologie*. Ed. by Martin-Dietrich Gleßgen and Franz Lebensanft. Tübingen: Max Niemeyer Verlag, 1–14.
- Godse, A. P. and D. A. Godse. 2007. *Fundamentals of Programming*. Pune: Technical Publications.
- Goffman, Erving. 1981. *Forms of Talk*. Oxford: Basil Blackwell.
- Goody, Jack. 1982. *Cooking, Cuisine and Class: a study in comparative sociology*. Cambridge: Cambridge University Press.
- Goody, Jack and Ian Watt. 1968. "The consequences of literacy". In: *Literacy in Traditional Societies*. Ed. by Jack Goody. Cambridge: Cambridge University Press, 27–84.
- Graver, Bruce and Ron Tetreault, eds. 2011. *Lyrical Ballads: an electronic scholarly edition*. University of Maryland. Available at: <http://www.rc.umd.edu/editions/LB/>. Accessed 21st May 2013.
- Grazia, Margreta de and Peter Stallybrass. 1993. "The Materiality of the Shakespearean Text". In: *Shakespeare Quarterly* 44, 255–83.
- Greetham, David C. 1987. "Challenges of Theory and Practice in the Editing of Hoccleve's *Regement of Princes*". In: *Manuscripts and Texts: Editorial Problems in Later Middle English Literature*. Ed. by Derek Pearsall. Cambridge: D. S. Brewer, 60–86.
- 1991. "The Manifestation and Accommodation of Theory in Textual Editing". In: *Devils and Angels: Textual Editing and Literary Theory*. Ed. by Philip Cohen. Charlottesville: University Press of Virginia, 78–102.
- 1992. Foreword. In: McGann, Jerome J. *A Critique of Modern Textual Criticism*. Charlottesville: University Press of Virginia, ix–xix.
- 1998. "Glosynge Is a Glorious Thyng, Certayn". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 287–302.
- 2004. "The Function of [Textual] Criticism at the Present Time". In: *Voice, Text, Hypertext: Emerging Practices in Textual Studies*. Ed. by Raimonda Modiano, Leroy F. Searle and Peter Shillingsburg. Seattle: University of Washington Press, 22–53.
- Greg, W. W. 1950–1951. "The Rationale of Copy-Text". In: *Studies in Bibliography* 3, 19–36.
- Grenier-Winther, Joan. 2004. "Server-Side Databases, the World Wide Web, and the Editing of Medieval Poetry: Tha Case of *La Belle dame qui eut mercy*". In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, 190–220.
- Grieco, Allen. 1992. "From the Cookbook to the Table: A Florentine Table and Italian Recipes of the Fourteenth and Fifteenth Centuries". In: *Du Manuscrit à la table: Essais sur la cuisine au Moyen Âge et répertoire des manuscrits médiévaux contenant des recettes culinaires*. Ed. by Carole Lambert. Montreal: Les Presses de l'Université de Montreal, 29–38.

- Grund, Peter. 2003. "The Golden Formulas: Genre Conventions of Alchemical Recipes in the Middle English Period". In: *Neuphilologische Mitteilungen* 104, 455–75.
- 2006. "Manuscripts as Sources for Linguistic Research: A Methodological Case Study Based on the *Mirror of Lights*". In: *Journal of English Linguistics* 34, 105–25.
- Grund, Peter, Merja Kytö and Matti Rissanen. 2004. "Editing the Salem witchcraft records: an exploration of a linguistic treasury". In: *American Speech* 79, 146–66.
- Gumbert, J. P. 2000. "Skins, Sheets and Quires". In: *New Directions in Later Medieval Manuscript Studies*. Ed. by Derek Pearsall. Rochester, NY: York Medieval Press, 81–90.
- Gumbrecht, Hans Ulrich. 1998. "Play Your Roles Tactfully! About the Pragmatics of Text-Editing, the Desire for Identification, and the Resistance to Theory". In: *Editing texts: texte edieren*. Ed. by Glenn Most. Göttingen: Vandenhoeck & Ruprecht, 237–50.
- Gunder, Anna. 2001. "Forming the Text, Performing the Work - Aspects of Media, Navigation, and Linking". In: *Human IT* 5. Available at: <http://etjanst.hb.se/bhs/ith/23-01/ag.htm>. Accessed 3rd January 2013.
- Guralnik, Benjamin. 2002. "Reverse Linking Explained". In: *Microsoft Developer Network Library*. Available at: <http://msdn.microsoft.com/en-us/library/ms950800.aspx>. Accessed 26th September 2013.
- Guy, Gregory. 2003. "Variationist Approaches to Phonological Change". In: *The Handbook of Historical Linguistics*. Ed. by Brian D. Joseph and Richard D. Janda. London: Blackwell Publishing, 369–400.
- Görlach, Manfred. 1992. "Text-types and language history: the cookery recipe". In: *History of Englishes. New Methods and Interpretations in Historical Linguistics*. Ed. by Matti Rissanen et al. Berlin: Mouton de Gruyter, 736–61.
- 2004. *Text Types and the History of English*. Berlin: Mouton de Gruyter.
- Hajo, Cathy Moran. 2010. "The sustainability of the scholarly edition in a digital world". In: *Proceedings of the International Symposium on XML for the Long Haul: Issues in the Long-term Preservation of XML*. Presented at International Symposium on XML for the Long Haul: Issues in the Long-term Preservation of XML, Montréal, Canada, August 2, 2010. Available at: <http://www.balisage.net/Proceedings/vol16/html/Hajo01/BalisageVol16-Hajo01.html>. Accessed 9th May 2012.
- Halliday, M. A. K. 1973. *Explorations in the Functions of Language*. London: Edward Arnold.
- 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- 2004. *An Introduction to Functional Grammar*. 3rd ed. London: Edward Arnold.
- Halliday, M. A. K. and Ruqaiya Hasan. 1976. *Cohesion in English*. 3rd ed. London: Longman.
- Halliday, M. A. K., Angus McIntosh and Peter Strevens. 1968. *The Linguistic Sciences and Language Teaching*. London: Longman.
- Halmari, Helena and Timothy Regetz. 2011. "Syntactic aspects of code-switching in Oxford, MS Bodley 649". In: *Code-switching in Early English*. Ed. by Herbert Schendl and Laura Wright. Berlin: De Gruyter Mouton, 115–53.

- Hamel, Mary. 1998. "The Use of Sources in Editing Middle English Texts". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 203–16.
- Hammond, P. W. 1993. *Food and Feast in Medieval England*. Stroud: Sutton Publishing.
- Hanna III, Ralph. 1986. "Booklets in medieval manuscripts: further considerations". In: *Studies in Bibliography* 39, 100–11.
- 1987. "Problems of 'Best Text' Editing and the Hengwrt Manuscript of *The Canterbury Tales*". In: *Manuscripts and Texts: Editorial Problems in Later Middle English Literature*. Ed. by Derek Pearsall. Cambridge: D. S. Brewer, 87–94.
  - 1992. "Producing Manuscripts and Editions". In: *Crux and Controversy in Middle English Textual Criticism*. Ed. by Alistair J. Minnis and Charlotte Brewer. Cambridge: D. S. Brewer, 109–30.
  - 2000. "Reconsidering the Auchinleck Manuscript". In: *New Directions in Later Medieval Manuscript Studies*. Ed. by Derek Pearsall. Rochester, NY: York Medieval Press, 91–102.
  - 2011. "Literacy, schooling, universities". In: *The Cambridge Companion to Medieval English Culture*. Ed. by Andrew Galloway. Cambridge: Cambridge University Press, 172–94.
- Hargreaves, Henry. 1981. "Some problems in indexing Middle English recipes". In: *Middle English Prose: Essays on Bibliographic Problems*. Ed. by A. S. G. Edwards and Derek Pearsall. New York: Garland Publishing, 91–113.
- BL Help: Harley MSS. "Harley Manuscripts". In: *British Library: Help for Researchers*. Available at: <http://www.bl.uk/reshelp/findhelpprestype/manuscripts/harleymss/harleymss.html>. Accessed 6th January 2014.
- Harris, Kate. 1989. "Patrons, buyers and owners: the evidence for ownership and the rôle of book owners in book production and the book trade". In: *Book Production and Publishing in Britain 1375-1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 163–99.
- Haspelmath, Martin. 2011. "The indeterminacy of word segmentation and the nature of morphology and syntax". In: *Folia Linguistica* 45, 31–80.
- Haugen, Odd Einar. 2004. "Parallel Views: Multi-level Encoding of Medieval Nordic Primary Sources". In: *Literary and Linguistic Computing* 19, 73–91.
- ed. 2008. *The Menota handbook: Guidelines for the electronic encoding of Medieval Nordic primary sources*. Version 2.0. Bergen: Medieval Nordic Text Archive. Available at: <http://www.menota.org/guidelines>. Accessed 5th January 2013.
  - ed. 2009. *MUFI character recommendation. Characters in the official Unicode Standard and in the Private Use Area for Medieval texts written in the Latin alphabet*. Version 3.0. Medieval Unicode Font Initiative (MUFI). Available at: <http://www.mufi.info/specs/>. Accessed 8th May 2012.
- HC. 1991. *The Helsinki Corpus of English Texts*. Helsinki: Department of Modern Languages, University of Helsinki. Compiled by Matti Rissanen (Project leader), Merja Kytö (Project secretary); Leena Kahlas-Tarkka, Matti Kilpiö (Old English); Saara Nevanlinna, Irma Taavitsainen (Middle English); Terttu Nevalainen, Helena Raumolin-Brunberg (Early Modern English).
- Heath, Shirley Brice. 1984. "Oral and literate traditions". In: *International Social Science Journal* 36, 41–57.

- Heawood, Edward. 1930. "Sources of early English paper-supply". In: *The Library* 10, 282–307.
- Hector, L. C. 1966. *The Handwriting of English Documents*. London: Edward Arnold (Publishers) Ltd.
- HC TEI XML. 2011. *Helsinki Corpus TEI XML Edition*. designer Terttu Nevalainen et al. implementer Henri Kauhanen and Ville Marttila. Based on The Helsinki Corpus of English Texts (1991). Helsinki. Available on request from The Research Unit for Variation, Contacts and Change in English (VARIENG).
- Henisch, Bridget Ann. 1978. *Fast and Feast: Food in Mediaeval Society*. University Park, PA: Pennsylvania State University Press.
- Henslow, George, ed. 1889. *Medical Works of the Fourteenth Century together with a List of Plants recorded in Contemporary Writings, with their Identifications*. London: Chapman and Hall Ltd.
- Herbermann, Charles G., ed. 1909. *The Catholic Encyclopedia: an international work of reference on the constitution, doctrine, discipline, and history of the Catholic Church*. Vol. 5. New York: Robert Appleton.
- Hettel, Jacqueline. 2013. "A Family Portrait of Computational Language Analysis". In: *A Linguist in the Library*. Blog entry. Available at: <http://www.linguabrararian.com/a-family-portrait-of-computational-language-analysis/>. Accessed 24th January 2014.
- Heyworth, P. L. 1989. *Letters of Humfrey Wanley, Palaeographer, Anglo-Saxonist, Librarian: 1672–1726*. Oxford: Clarendon Press.
- Hieatt, Constance B., ed. 1988. *An ordinance of pottage: an edition of the fifteenth century culinary recipes in Yale University's MS Beinecke 163. with a commentary on the recipes and adapted versions for the modern kitchen*. illustrator Eulalia Pensado. London: Prospect Books.
- 1992. "Listing and analysing the medieval English culinary recipe collections: a project and its problems". In: *Du Manuscrit à la table: Essais sur la cuisine au Moyen Âge et répertoire des manuscrits médiévaux contenant des recettes culinaires*. Ed. by Carole Lambert. Montreal: Les Presses de l'Université de Montréal, 15–21.
  - 1996. "The Middle English culinary recipes in MS Harley 5401: an edition and commentary". In: *Medium Ævum* 65, 54–71.
  - 1998a. "Editing Middle English Culinary Manuscripts". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 133–40.
  - 1998b. "Making Sense of Medieval Culinary Records: Much Done, But Much More to Do". In: *Food and Eating in Medieval Europe*. Ed. by Martha Carlin and Joel Rosenthal. London: The Hambledon Press, 101–16.
  - 2004. "The third fifteenth-century cookery book: a newly identified group within a family". In: *Medium Ævum* 73, 27–42.
  - 2008. *A Gathering of Medieval English Recipes*. Textes Vernaculaire du Moyen Âge 5. Turnhout: Brepols.
  - 2012. *Cocatrice and Lampray Hay: Late Fifteenth-century Recipes from Corpus Christi College Oxford*. Totnes: Prospect Books.
- Hieatt, Constance B. and Sharon Butler, eds. 1976. *Plen Delit: Medieval Cookery for Modern Cooks*. Toronto: University of Toronto Press.

- Hieatt, Constance B. and Sharon Butler, eds. 1985. *Curie on Inglisch: English Culinary Manuscripts of the Fourteenth Century (including the Forme of Cury)*. Early English Text Society, Supplementary Series 8. Oxford: Oxford University Press.
- Hieatt, Constance B. and Rudolf Grewe, eds. and trans. 2001. *Libellus de arte coquinaria. An Early Northern Cookery Book*. Tempe, AZ: Arizona Center for Medieval and Renaissance Studies.
- Hieatt, Constance B. and Robin F. Jones. 1986. "Two Anglo-Norman Culinary Collections Edited from British Library Manuscripts Additional 32085 and Royal 12.C.xii". In: *Speculum* 61, 859–82.
- Hieatt, Constance B., Carole Lambert et al. 1992. "Répertoire des Manuscrits Médiévaux Contenant des Recettes Culinaires". In: *Du Manuscrit à la table: Essais sur la cuisine au Moyen Âge et répertoire des manuscrits médiévaux contenant des recettes culinaires*. Ed. by Carole Lambert. Montreal: Les Presses de l'Université de Montreal, 317–62.
- Hieatt, Constance B. and Terry Nutter. 2006. *Concordance of English Recipes: Thirteenth through Fifteenth Centuries*. Medieval & Renaissance Texts & Studies 312. Tempe, AZ: Arizona Centre for Medieval and Renaissance Studies.
- Hiippala, Tuomo. 2013. "Modelling the structure of a multimodal artefact". PhD thesis. Helsinki: University of Helsinki. Available at: <http://urn.fi/URN:ISBN:978-952-10-9427-9>. Accessed 24th January 2014.
- Hockey, Susan. 2004. "The Reality of Electronic Editions". In: *Voice, Text, Hypertext: Emerging Practices in Textual Studies*. Ed. by Raimonda Modiano, Leroy F. Searle and Peter Shillingsburg. Seattle: University of Washington Press, 361–77.
- Hodgett, Gerald A. J., ed., trans. and forew. 1972. *Stere Htt Wele: A Book of Medieval Refinements, Recipes and Remedies, from a Manuscript in Samuel Pepys's Library*. With an intro. by Della Smith. London: Cornmarket Press.
- Hoey, Michael. 1986. "The Discourse Colony: A Preliminary Study of a Neglected Discourse Type". In: *Talking About Text: Studies Presented to David Brazil on his Retirement*. Ed. by Malcolm Coulthard. Birmingham: English Language Research, University of Birmingham, 1–26.
- 2001. *Textual Interaction: An introduction to written discourse analysis*. London: Routledge.
- LC. 1999. "The Lampeter Corpus of Early Modern English Tracts". In: *ICAME Collection of English Language Corpora (CD-ROM)*. Ed. by Knut Hofland, Anne Lindebjerg and Jørn Thunestvedt. Comp. by Josef Schmied, Claudia Claridge and Rainer Siemund. Bergen: The HIT Centre, University of Bergen.
- Holinshed, Raphael. 1577. *The Laste volume of the Chronicles of England, Scotlande and Irelande, with their descriptions, conteyning The Chronicles of Englande from William Conquerour vntill this present tyme*. London: John Harrison. Facsimile copy provided by *Early English Books Online (EEBO)*.
- Honkapohja, Alpo, Samuli Kaislaniemi and Ville Marttila. 2009. "Digital Editions for Corpus Linguistics: Representing manuscript reality in electronic corpora". In: *Corpora: Pragmatics and Discourse. Papers from the 29th International Conference on English Language Research on Computerized Corpora (ICAME 29)*. Ascona, Switzerland, 14–18 May 2008. Ed. by Andreas H. Jucker, Daniel Schreier and Marianne Hundt. Amsterdam/New York: Rodopi.
- Hope, Jonathan. 2000. "Rats, bats, sparrows and dogs: biology, linguistics and the nature of Standard English". In: *The Development of Standard English 1300–1800*:

- Theories, Descriptions, Conflicts*. Ed. by Laura Wright. Cambridge: Cambridge University Press, 49–56.
- Horner, Patrick J. 2006. *A Macaronic Sermon Collection from Late Medieval England: Oxford MS Bodley 649*. Toronto: Pontifical Institute of Medieval Studies.
- Howard-Hill, T. H. 1991. "Variety in Editing and Reading: A Response to McGann and Shillingsburg". In: *Devils and Angels: Textual Editing and Literary Theory*. Ed. by Philip Cohen. Charlottesville: University Press of Virginia, 44–56.
- Huddleston, Rodney D. and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Huitfeldt, Claus. 2006. "Philosophy Case Study". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 181–96.
- Huitfeldt, Claus and C. M. Sperberg-McQueen. 2004. "What is transcription?" In: *Literary and Linguistic Computing* 19, 295–310.
- Hunston, Susan. 2008. "Collection strategies and design decisions". In: *Corpus Linguistics: An International Handbook*. Ed. by Anke Lüdeling and Merja Kytö. Berlin: Walter de Gruyter, 154–68.
- Hunt, Tony. 1990. *Popular medicine in thirteenth-century England: introduction and texts*. Cambridge: Boydell & Brewer Ltd.
- 2000. "Code-switching in medical texts". In: *Multilingualism in Later Medieval Britain*. Ed. by David Trotter. Cambridge: D. S. Brewer, 131–48.
- Hunter, Michael et al. 1983. *Elias Ashmole 1617-1692 - the founder of the Ashmolean Museum and his world: A Tercentenary Exhibition (27 April to 31 July 1983)*. Oxford: Ashmolean Museum.
- Hussey, S. S. 1992. "Editing *The Scale of Perfection*: Return to Recension". In: *Crux and Controversy in Middle English Textual Criticism*. Ed. by Alistair J. Minnis and Charlotte Brewer. Cambridge: D. S. Brewer, 97–107.
- Iacob, Ionut E. and Alex Dekhtyar. 2005. "Processing XML documents with overlapping hierarchies". In: *JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*. New York, p. 409.
- Iglesias-Rábade, Luis. 1996. "The multi-lingual pulpit in England (1100–1500)". In: *Neophilologus* 80, 479–92.
- Ingham, Richard. 2010a. "Anglo-Norman: New Themes, New Contexts". In: *The Anglo-Norman Language and its Contexts*. Ed. by Richard Ingham. York: York Medieval Press, 1–7.
- 2010b. "Later Anglo-Norman as a Contact Variety of French?" In: *The Anglo-Norman Language and its Contexts*. Ed. by Richard Ingham. York: York Medieval Press, 8–25.
- 2011. "Code-switching in later medieval English lay subsidy rolls". In: *Code-switching in Early English*. Ed. by Herbert Schendl and Laura Wright. Berlin: De Gruyter Mouton, 95–114.
- Jackson, H. J. 2004. "Editing and Auditing Marginalia". In: *Voice, Text, Hypertext: Emerging Practices in Textual Studies*. Ed. by Raimonda Modiano, Leroy F. Searle and Peter Shillingsburg. Seattle: University of Washington Press, 72–80.
- Jacobs, Nicolas. 1998. "Kindly Light or Foxfire? The Authorial Text Reconsidered". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 3–14.

- Janda, Richard D. and Brian D. Joseph. 2003. "On Language, Change, and Language Change — Or, Of History, Linguistics, and Historical Linguistics". In: *The Handbook of Historical Linguistics*. Ed. by Brian D. Joseph and Richard D. Janda. London: Blackwell Publishing, 3–180.
- Jaro, Matthew A. 1989. "Advances in record linkage methodology as applied to the 1985 census of Tampa Florida". In: *Journal of the American Statistical Association* 84, 414–20.
- Jefferson, Lisa. 2000. "The language and vocabulary of the fourteenth- and early fifteenth-century records of the Goldsmiths' Company". In: *Multilingualism in Later Medieval Britain*. Ed. by David Trotter. Cambridge: D. S. Brewer, 175–212.
- Jensen, Michael. 2004. "Intermediation and its Malcontents: Validating Professionalism in the Age of Raw Dissemination". In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, Ray Siemens and John Unsworth. Malden, MA: Blackwell Publishing, 543–56.
- Jespersen, Otto. 1924. *The philosophy of grammar*. London: George Allen & Unwin Ltd.
- Johansson, Karl G. 2004. "Computing Medieval Primary Sources from the Vadstena Monastery: Arguments for the Primary Source Text". In: *Literary and Linguistic Computing* 19, 93–104.
- Johansson, Stig. 2008. "Some aspects of the development of corpus linguistics in the 1970s and 1980s". In: *Corpus Linguistics: An International Handbook*. Ed. by Anke Lüdeling and Merja Kytö. Berlin: Walter de Gruyter, 33–53.
- Johansson, Stig, Geoffrey Leech and Helen Goodluck. 1978. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Oslo: Department of English, University of Oslo. Available at: <http://khnt.hit.uib.no/icame/manuals/lob/INDEX.HTM>. Accessed 25th November 2013.
- Johnes, Thomas, ed. 1806. *Sir John Froissart's Chronicles of England, France, Spain, and the adjoining countries, from the latter part of Edward II. to the coronation of Henry IV. Newly translated from the French editions, with variations and additions from many celebrated MSS.* 2nd ed. Vol. XII. London: Longman, Hurst, Rees and Orme & J. White.
- Jones, Claire. 1998. "Formula and formulation: 'efficacy phrases' in medieval English medical manuscripts". In: *Neuphilologische Mitteilungen* 99, 199–209.
- 2000. "Vernacular literacy in late-medieval England: the example of East Anglian medical manuscripts". PhD thesis. University of Glasgow. Available at: <http://theses.gla.ac.uk/633/>. Accessed 9th August 2013.
- Joyce, James. 1984. *Ulysses. A Critical and Synoptic Edition*. Ed. by Hans Walter Gabler, Wolfhard Steppe and Claus Melchior. New York: Garland Publishers.
- Kane, George, ed. 1960. *Piers Plowman. The A version: Will's visions of Piers Plowman and do-well. an edition in the form of Trinity College, Cambridge Ms.R.3.14 corrected from other manuscripts, with variant readings*. London: The Athlone Press.
- Kane, George and Ethelbert Talbot Donaldson, eds. 1975. *Piers Plowman. The B version: Will's visions of Piers Plowman, do-well, do-better and do-best. an edition in the form of Trinity College, Cambridge Ms.B.15.17, corrected and restored from the known evidence, with variant readings*. London: The Athlone Press.
- Keen, Maurice. 1990. *English Society in the Later Middle Ages 1348–1500*. London: Penguin Books.



- Keene, Derek. 2000. "Metropolitan values: migration, mobility and cultural norms, London 1100–1700". In: *The Development of Standard English 1300–1800: Theories, Descriptions, Conflicts*. Ed. by Laura Wright. Cambridge: Cambridge University Press, 93–114.
- Keiser, George R. 1988. "Review: *Curye on Inglysch*". In: *Speculum* 63, 410–2.
- 1998a. *A manual of the writings in Middle English 1050-1500. Vol. 10: Works of science and information*. New Haven: Connecticut Academy of Arts and Sciences.
- 1998b. "Editing Scientific and Practical Writings". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 109–22.
- 1998c. *A manual of the writings in Middle English 1050-1500. Vol. 10: Works of science and information*. Ed. by Albert E. Hartung. New Haven: Connecticut Academy of Arts and Sciences.
- Keller, Henning. 1971. "Die me. Rezepte des Ms. Harley 2253". In: *Archiv für das Studium der neueren Sprachen und Literaturen* 207, 94–100.
- Kerby-Fulton, Kathryn. 2000. "Professional Readers of Langland at Home and Abroad: New Directions in the Political and Bureaucratic Codicology of *Piers Plowman*". In: *New Directions in Later Medieval Manuscript Studies*. Ed. by Derek Pearsall. Rochester, NY: York Medieval Press, 103–29.
- Kiernan, Kevin. 2006. "Digital facsimiles in editing". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 262–8.
- Kingsford, Charles L., ed. 1962. *Prejudice and promise in fifteenth century England*. London: F. Cass.
- Kirschenbaum, Matthew G. 2002. "Editing the Interface: Textual Studies and First Generation Electronic Objects". In: *TEXT: Transactions of the Society for Textual Scholarship* 14, 15–51.
- 2004. "'So the colors Cover the Wires': Interface, Aesthetics, and Usability". In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, Ray Siemens and John Unsworth. Malden, MA: Blackwell Publishing, 523–42.
- Klinck, Anne L. 2004. "Editing *Cursor Mundi*: Stemmata and the 'Open' Text". In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, 3–13.
- Kline, Mary-Jo and Susan Holbrook Perdue. 2008. *A Guide to Documentary Editing*. Charlottesville: University of Virginia Press.
- Knight, Leah. 2009. *Of Books and Botany in Early Modern England: Sixteenth-Century Plants and Print Culture*. Farnham: Ashgate.
- Kress, Gunther and Terry Threadgold. 1988. "Towards a Social Theory of Genre". In: *Southern Review* 21, 215–43.
- Kroch, Anthony, Beatrice Santorini and Lauren Delfs. 2004. *Penn-Helsinki Parsed Corpus of Early Modern English*. Philadelphia. Available on CD-ROM by request at <http://www.ling.upenn.edu/hist-corpora/PPCME-RELEASE-2/index.html>.
- Kroch, Anthony and Ann Taylor. 2000. *Penn-Helsinki Parsed Corpus of Middle English, second edition*. Philadelphia. Available on CD-ROM by request at <http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-3/index.html>.

- Kytö, Merja, ed. 1996. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts*. 3rd ed. Helsinki: Department of English, University of Helsinki.
- Kytö, Merja, Peter Grund and Terry Walker. 2007. "Regional variation and the language of English witness depositions 1560-1760: constructing a 'linguistic' edition in electronic form". In: *Towards Multimedia in Corpus Studies*. Ed. by Päivi Pahta et al. Helsinki: Research Unit for Variation, Contacts and Change in English (VARIENG), University of Helsinki. Available at: [http://www.helsinki.fi/varieng/journal/volumes/02/kyto\\_et\\_al/](http://www.helsinki.fi/varieng/journal/volumes/02/kyto_et_al/). Accessed 5th December 2012.
- Kytö, Merja and Terry Walker. 2003. "The linguistic study of Early Modern English speech-related texts: How 'bad' can 'bad' data be?" In: *Journal of English Linguistics* 31, 221-48.
- Kytö, Merja, Terry Walker and Peter Grund. 2007. "English witness depositions 1560-1760: An electronic text edition". In: *ICAME Journal* 31, 65-86. Available at: <http://icame.uib.no/ij24/merja.pdf>. Accessed 8th December 2012.
- Laing, Margaret. 1989. "Dialectal Analysis and Linguistically Composite Texts in Middle English". In: *Middle English Dialectology: Essays on Some Principles and Problems*. Ed. by Margaret Laing. Aberdeen: Aberdeen University Press, 150-69.
- 2000. "Never the twain shall meet: Early Middle English—the East-West divide". In: *Placing Middle English in Context*. Ed. by Terttu Nevalainen et al. Berlin: Mouton de Gruyter, 97-124.
- Lambert, Carole. 1989. "Trois réceptaires culinaires médiévaux: les Enseignementz, les Doctrines et le Modus. Edition critique et glossaire détaillé". PhD thesis. Université de Montréal.
- Landow, George P. 1992. *Hypertext: The Convergence of Contemporary Critical Theory and Technology*. Baltimore: The Johns Hopkins University Press.
- 1997. *Hypertext 2.0. Being a revised, amplified edition of Hypertext: The Convergence of Contemporary Critical Theory and Technology*. Baltimore: The Johns Hopkins University Press.
- Langacker, Ronald W. 1972. *Fundamentals of linguistic analysis*. New York: Harcourt Brace Jovanovich.
- Lass, Roger. 1987. *The Shape of English*. London: J. M. Dent & Sons.
- ed. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.
- 2004a. "Texts as linguistic objects". In: *Categorization in the history of English*. Ed. by Christian J. Kay and Jeremy J. Smith. Amsterdam: John Benjamins, 147-57.
- 2004b. "Ut custodiant litteras: Editions, Corpora and Witnesshood". In: *Methods and Data in English Historical Dialectology*. Ed. by Maria Dossena and Roger Lass. Bern: Peter Lang, 21-48.
- Laurioux, Bruno. 2002. *Manger au Moyen Âge*. Paris: Hachette Littératures.
- Lavagnino, John. 1995. "Reading, Scholarship, and Hypertext Editions". In: *TEXT: Transactions of the Society for Textual Scholarship* 8, 109-24.
- 2006. "When not to use TEI". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 334-8.

- 2009. "Access". In: *Literary and Linguistic Computing* 24, 63–76.
- Leckie-Tarry, Helen. 1993. "The specification of a text: register, genre and language teaching". In: *Register Analysis: Theory and Practice*. Ed. by Mohsen Ghadessy. London: Pinter Publishers, 26–42.
- 1995. *Language and Context: A functional linguistic theory of register*. Ed. by David Birch. London: Pinter Publishers.
- Lee, Benny. 2001. "Mutual knowledge, background knowledge and shared beliefs: Their roles in establishing common ground". In: *Journal of Pragmatics* 33, 21–44.
- Leech, Geoffrey. 2005. "Adding Linguistic Annotation". In: *Developing Linguistic Corpora: a Guide to Good Practice*. Ed. by Martin Wynne. Oxford: Oxbow Books, 17–29. Available at: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter2.htm>. Accessed 29th January 2014.
- Lehmann, Gilly. 2003. *The British Housewife: Cookery Books, Cooking and Society in Eighteenth-century Britain*. Totnes: Prospect Books.
- Lehmberg, Timm and Kai Wörner. 2008. "Annotation standards". In: *Corpus Linguistics: An International Handbook*. Ed. by Anke Lüdeling and Merja Kytö. Berlin: Walter de Gruyter, 484–501.
- Leslie, Michael. 1993. "Electronic Editions and the Hierarchy of Texts". In: *The Politics of the Electronic Text*. Ed. by Warren Chernaik, Caroline Davis and Marilyn Deegan. Oxford: Office for Humanities Communication, Oxford University Computing Services, 41–52.
- Levenshtein, Vladimir I. 1966. "Binary codes capable of correcting deletions, insertions, and reversals. Proceedings of the Academy of Sciences of USSR, Physics section". In: *Soviet physics - Doklady* 10. Translated from Doklady Akademii Nauk SSSR, Vol. 163, No. 4, pp. 845–8, August 1965., 707–10.
- Lewis, Charlton T. and Charles Short, eds. 1879. *A Latin Dictionary. Founded on Andrews' edition of Freund's Latin dictionary, revised, enlarged, and in great part rewritten*. Oxford: Clarendon Press.
- Longnon, Jean and Raymond Cazelles, eds. and introd. 1993. *Les Très Riches Heures du duc de Berry*. London: Thames and Hudson.
- Lowe, Kathryn. 1998. "Lay literacy in Anglo-Saxon England and the development of the chirograph". In: *Anglo-Saxon Manuscripts and their Heritage*. Ed. by Phillip Pulsiano and Elaine M. Treharne. Aldershot: Ashgate, 161–204.
- Lucas, Peter J. 1998. "The Treatment of Language". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 169–83.
- Lyall, R. J. 1989a. "Books and book owners in fifteenth-century Scotland". In: *Book Production and Publishing in Britain 1375-1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 239–56.
- 1989b. "Materials: the paper revolution". In: *Book Production and Publishing in Britain 1375-1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 11–29.
- Maas, Paul. 1958. *Textual criticism*. Trans. by Barbara Flower. London: Clarendon Press.
- MacCracken, H. N., ed. 1934. *The minor poems of John Lydgate, edited from all available mss., with an attempt to establish the Lydgate canon*. Vol. 2: *Secular*

- poems*. Early English Text Society, Original Series 192. London: Oxford University Press.
- Machan, Tim William. 1992. "Middle English Text Production and Modern Textual Criticism". In: *Crux and Controversy in Middle English Textual Criticism*. Ed. by Alistair J. Minnis and Charlotte Brewer. Cambridge: D. S. Brewer, 1–18.
- 1994. *Textual Criticism and Middle English Texts*. Charlottesville: University Press of Virginia.
- 2009. "French, English, and the Late Medieval Linguistic Repertoire". In: *Language and Culture in Medieval Britain: the French of England c. 1100-1500*. Ed. by Jocelyn Wogan-Browne et al. York: York Medieval Press, 363–72.
- Madan, Falconer, H.H.E. Craster and N.D. Young. 1937. *A summary catalogue of western manuscripts in the Bodleian Library at Oxford which have not hitherto been catalogued in the Quarto series. Collections and miscellaneous mss. acquired during the second half of the 17th century, nos. 3491-8716*. Vol. 2.2. Oxford: Clarendon Press.
- Madison, Olivia et al., eds. 2009. *Functional Requirements for Bibliographic Records*. Final Report. International Federation of Library Associations and Institutions. Available at: <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>. Accessed 4th December 2012.
- Mahoney, Anne. 2006. "Epigraphy". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 224–37.
- Malhotra, Ashok and Paul V. Biron, eds. 2004. *XML Schema Part 2: Datatypes*. 2nd ed. W3C Recommendation. The World Wide Web Consortium. Available at: <http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/>. Accessed 24th February 2012.
- Malinowski, Bronislaw. 1923. "The Problem of Meaning in Primitive Languages". In: *The Meaning of Meaning*. Ed. by C. K. Ogden and I. A. Richards. New York: Harcourt, Brace and World, 296–336.
- 1935. *Coral gardens and their magic: a study of the methods of tilling the soil and of agricultural rites in the Trobriand Islands, Vol. 2: The language of magic and gardening*. London: George Allen & Unwin.
- Manly, John M. and Edith Rickert, eds. 1940. *The text of the Canterbury Tales studied on the basis of all known manuscripts*. Chicago: University of Chicago Press.
- Manzalaoui, Mahmoud. 1974. "Chaucer and Science". In: *Geoffrey Chaucer*. Ed. by Derek Brewer. London: Bell, 224–61.
- ed. 1977. *Secretum Secretorum: nine English versions*. Early English Text Society, Original Series 276. Oxford: Oxford University Press.
- Mapstone, Sally. 1997. "Kingship and the *Kingis Quair*". In: *The Long Fifteenth Century: Essays for Douglas Gray*. Ed. by Helen Cooper and Sally Mapstone. Oxford: Clarendon Press, 51–69.
- Marcoux, Yves. 2008. "Graph characterization of overlap-only TexMECS and other overlapping markup formalisms". In: *Proceedings of Balisage: The Markup Conference 2008*. Presented at Balisage: The Markup Conference 2008, Montréal, Canada, August 12 - 15, 2008. Available at: <http://www.balisage.net/Proceedings/vol11/html/Marcoux01/BalisageVol11-Marcoux01.html>. Accessed 9th May 2012.
- Markus, Manfred. 1997. "Normalization of Middle English prose: possibilities and limits. Papers from the seventeenth International Conference on English Lan-

- guage Research on Computerized Corpora (ICAME 17), Stockholm, May 15-19, 1996". In: *Corpus-based Studies in English*. Ed. by Magnus Ljung. Amsterdam: Rodopi, 211-26.
- 2000. "Normalising the Word Forms in the *Ayenbite of Inwit*". In: *Placing Middle English in Context*. Ed. by Terttu Nevalainen et al. Berlin: Mouton de Gruyter, 181-98.
- Martens, Gunter. 1995. "What Is a Text? Attempts at Defining a Central Concept in Editorial Theory". In: *Contemporary German Editorial Theory*. Ed. by Hans Walter Gabler, George Bornstein and Gillian Borland Pierce. Ann Arbor: University of Michigan Press, 209-31.
- Martin, J. R. 1992. *English Text: System and Structure*. Amsterdam: John Benjamins Publishing.
- Marttila, Ville. 2006. "Recipes for *Potage Dyvers*: A contextualised edition of the culinary recipe collection in Durham University Library MS Cosin V.III.11 ff. 1-21". MA thesis. University of Helsinki.
- 2009. "Mincing Words: A Diachronic View on English Cutting Verbs". In: *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX 2)*. Ed. by Jukka Tyrkkö, Rod McConchie and Alpo Honkapohja. Somerville, Mass.: Cascadilla Proceedings Project, 104-22. Available at: <http://www.lingref.com/cpp/hel-lex/2008/paper2171.pdf>. Accessed 17th October 2013.
- 2011. "New arguments for new audiences: a corpus-based analysis of interpersonal strategies in early modern English medical recipes". In: *Medical Writing in Early Modern English*. Ed. by Irma Taavitsainen and Päivi Pahta. Cambridge: Cambridge University Press, 135-57.
- Forthcoming. *Recipes for Potage Dyvers: A contextualised edition of the culinary recipe collection in Durham University Library MS Cosin V.III.11 ff. 1-21*. Textes Vernaculaire du Moyen Âge 13. Turnhout: Brepols.
- Marvin, Julia. 2004. "The Unassuming Reader: F. W. Maitland and the Editing of Anglo-Norman". In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, 14-36.
- Matheson, Lister M. 1999. *Death and Dissent: two fifteenth-century chronicles*. Woodbridge, Suffolk: Boydell & Brewer Ltd.
- Mathiessen, Christian. 1993. "Register in the round: diversity in a unified theory of register analysis". In: *Register Analysis: Theory and Practice*. Ed. by Mohsen Ghadessy. London: Pinter Publishers, 221-92.
- Matthews, Peter H. 1991. *Morphology*. 2nd ed. Cambridge: Cambridge University Press.
- McCarren, Vincent. 1998. "Editing Glossographical Text: To Marrow and to Marrow and to Marrow". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 141-55.
- McCarren, Vincent and Douglas Moffat, eds. 1998. *A Guide to Editing Middle English*. Ann Arbor: University of Michigan Press.
- McDermott, Anne and Marcus Walsh. 1991. "Editing Johnson's *Dictionary*: some editorial and textual considerations. Essays in Honour of James T. Boulton". In: *The Theory and Practice of Text-Editing*. Ed. by Ian Small and Marcus Walsh. Cambridge: Cambridge University Press, 35-61.

- McGann, Jerome J. 1985. "The Monks and the Giants: Textual and Bibliographical Studies and the Interpretation of Literary Works". In: *Textual Criticism and Literary Interpretation*. Ed. by Jerome J. McGann. Chicago: The University of Chicago Press, 180–99.
- 1991. *The Textual Condition*. Princeton: Princeton University Press.
- 1992. *A Critique of Modern Textual Criticism*. 2nd ed. Charlottesville: University Press of Virginia.
- 2004. "Imagining What You Don't Know: The Theoretical Goals of the Rossetti Archive". In: *Voice, Text, Hypertext: Emerging Practices in Textual Studies*. Ed. by Raimonda Modiano, Leroy F. Searle and Peter Shillingsburg. Seattle: University of Washington Press, 378–97.
- McGillivray, Murray. 1994. "Towards a Post-Critical Edition: Theory, Hypertext, and the Presentation of Middle English Works". In: *TEXT: Transactions of the Society for Textual Scholarship* 7, 175–99.
- McIntosh, Angus. 1989a. "Some Notes on the Language and Textual Transmission of the *Scottish Troy Book*". In: *Middle English Dialectology: Essays on Some Principles and Problems*. Ed. by Margaret Laing. Aberdeen: Aberdeen University Press, 237–55.
- 1989b. "Word Geography in the Lexicography of Medieval English". In: *Middle English Dialectology: Essays on Some Principles and Problems*. Ed. by Margaret Laing. Aberdeen: Aberdeen University Press, 86–97.
- McIntosh, Angus, M. L. Samuels and Michael Benskin. 1986. *A Linguistic Atlas of Late Mediaeval English*. 4 vols. Aberdeen: Aberdeen University Press.
- McKitterick, Rosamond, ed. 1990. *The Uses of Literacy in Early Medieval Europe*. Cambridge: Cambridge University Press.
- McLoughlin, Tim. 2008. "Bridging the Gap". In: *Jahrbuch für Computerphilologie* 10. Ed. by Georg Braungart, Peter Gendolla and Fotis Jannidis. Available at: <http://computerphilologie.de/jg08/mclough.pdf>. Accessed 9th January 2013.
- McLuhan, Marshall. 1962. *The Gutenberg galaxy: the making of typographic man*. London: Routledge.
- Mead, William Edward. 1931. *The English Medieval Feast*. London: George Allen & Unwin Ltd.
- Meale, Carol M. 1989. "Patrons, buyers and owners: book production and social status". In: *Book Production and Publishing in Britain 1375–1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 201–38.
- Meecham-Jones, Simon. 2011. "'Gadryng Tokedre of Medecyne in the Partye of Cyrugie': Strategies of code-switching in the Middle English translations of Chauliac's *Chirurgia Magna*". In: *Code-switching in Early English*. Ed. by Herbert Schendl and Laura Wright. Berlin: De Gruyter Mouton, 253–80.
- Mennell, Stephen. 1986. *All manners of food: eating and taste in England and France from the Middle Ages to the present*. New York: Blackwell.
- Mennell, Stephen, Anne Murcott and Anneke H. van Otterloo. 1992. *The sociology of food: eating, diet and culture*. Longon: Sage Publishing.
- Meredith, Peter. 2004. "The Language of Medieval Cookery". In: *The English Cookery Book: Historical Essays*. Ed. by Eileen White. Totnes: Prospect Books, 28–54.
- Meurman-Solin, Anneli. 2001. "Structured Text Corpora in the Study of Language Variation and Change". In: *Literary and Linguistic Computing* 16, 5–27.

- 2007. *Manual to the Corpus of Scottish Correspondence (CSC)*. Helsinki: University of Helsinki. Available at: <http://www.helsinki.fi/varieng/csc/manual/>. Accessed 22nd April 2013.
- Middle English Dictionary*. 2001. Ann Arbor: University of Michigan. Available at: <http://quod.lib.umich.edu/m/med/>. Accessed 9th May 2012.
- Milham, Mary Ella, ed. 1998. *Platina: On Right Pleasure and Good Health*. Tempe, Arizona: Medieval & Renaissance Texts & Studies.
- Miller, Gary D. 1997. "The Morphological Legacy of French: Borrowed Suffixes on Native Bases in Middle English". In: *Diachronica* 14, 233–64.
- Mills, Maldwyn. 1998. "Using the *Linguistic Atlas of Late Mediaeval English*". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 185–202.
- Milroy, James. 1992. *Linguistic Variation and Change*. Oxford: Blackwell.
- Minnis, Alistair J., ed. 1984. *Medieval Theory of Authorship: Scholastic literary attitudes in the later Middle Ages*. London: Scolar Press.
- Minnis, Alistair J. and Charlotte Brewer, eds. 1992. *Crux and Controversy in Middle English Textual Criticism*. Cambridge: D. S. Brewer.
- Modiano, Raimonda, Leroy F. Searle and Peter Shillingsburg, eds. 2004. *Voice, Text, Hypertext: Emerging Practices in Textual Studies*. Seattle: University of Washington Press.
- Moffat, Douglas. 1998. "A Bibliographical Essay on Editing Methods and Authorial and Scribal Intention". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 25–57.
- Moffat, Douglas and Vincent McCarren. 1998. "A Bibliographical Essay on Editing Methods and Authorial and Scribal Intention". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 25–57.
- Mooney, Linne R., ed. 1995. *The Index of Middle English Prose, Handlist XI: The Library of Trinity College, Cambridge*. Cambridge: Boydell & Brewer.
- 1998. "Editing Astrological and Prognostic Texts". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 123–32.
- Morris, Richard, ed. 1862. *Liber Cure Cocorum: copied and edited from the Sloane MS. 1986*. Berlin: A. Asher & Co. Available at: <http://archive.org/details/libercurecocorum00morr>. Accessed 22nd October 2012.
- Mosin, Vladimir A. and Seid M. Traljic. 1957. *Filigranes des XIIIe et XIVe siècles*. 2 vols. Zagreb: Académie yougoslave des Sciences et des Beaux-Arts. Institut d'histoire.
- Mueller, Martin. 2009. "NUPOS: A part of speech tag set for written English from Chaucer to the present". Available at: <http://panini.northwestern.edu/mmuellder/nupos.pdf>. Accessed 25th January 2012.
- Mulon, Marianne. 1971. "Deux traités inédits d'art culinaire médiéval". In: *Bulletin philologique et historique jusqu'à 1610 du Comité des travaux historiques et scientifiques: Année 1968: Actes du 93<sup>e</sup> Congrès national des Sociétés savantes tenu à Tours. Volume 1: Problèmes de l'alimentation*, 396–435.
- Mussell, James and Suzanne Paylor. 2009. "Editions and Archives: Textual Editing and the Nineteenth-century Serials Edition (ncse)". In: *Text Editing. Print and*

- the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 137–57.
- Myers, A. R., ed. 1959. *The Household of Edward IV: The 'Black Book' and the Ordinance of 1478*. Manchester: Manchester University Press.
- Mäkinen, Martti. 2004. "Herbal Recipes and Recipes in Herbals—Intertextuality in Early English Herbals and Other Medical Writing". In: *Medical and Scientific Writing in Late Medieval English*. Ed. by Irma Taavitsainen and Päivi Pahta. Cambridge: Cambridge University Press, 144–73.
- Möhren, Frankwalt. 2000. "Onefold lexicography for a manifold problem?" In: *Multilingualism in Later Medieval Britain*. Ed. by David Trotter. Cambridge: D. S. Brewer, 157–68.
- Napier, Robina. 1882. *A Noble boke off cookry ffor a prynce houssolde or eny other estately houssolde: reprinted verbatim from a rare ms. in the Holkham collection*. 2nd ed. London: Elliot Stock.
- Nevalainen, Terttu and Ingrid Tieken-Boon van Ostade. 2006. "Standardisation". In: *A History of the English Language*. Ed. by Richard Hogg and David Denison. Cambridge: Cambridge University Press, 271–311.
- Nicholls, Jonathan. 1985. *The Matter of Courtesy: Medieval Courtesy Books and the Gawain-Poet*. Woodbridge: D. S. Brewer.
- Nichols, John, ed. 1790. *A collection of ordinances and regulations for the government of the Royal household, made in divers reigns. From King Edward III. to King William and Queen Mary. Also receipts in ancient cookery*. London: Society of Antiquaries.
- Nichols, Stephen G. 1986. "Introduction: Philology in a Manuscript Culture". In: *Speculum* 61, 1–10.
- 1991. "On the Sociology of Medieval Manuscript Annotation". In: *Annotation and Its Texts*. Ed. by Stephen A. Barney. New York: Oxford University Press, 43–73.
- Nurmi, Arja. 1999. "The Corpus of Early English Correspondence Sampler (CEECS)". In: *ICAME Journal* 23, 53–64.
- Odgen, Margaret Sinclair, ed. 1938. *The Liber de diversis medicinis in the Thornton manuscript (MS. Lincoln Cathedral A.5.2)*. Early English Text Society, Original Series 207. London: Oxford University Press.
- Ong, Walter J. 1982. *Orality and Literacy: The Technologizing of the Word*. London: Routledge.
- Ore, Espen S. 1999. "Elektronisk Publisering: forskjellige utgaveformer og forgoldet til grunntekst(er) og endelige text(er). bidrag till en konferens anordnad av Nordisk nätverk för editionsfilologer 16-18 oktober 1998". In: *Vid texternas vägsål: textkritiska uppsatser*. Ed. by Lars Burman and Barbro Ståhle Sjönell. Stockholm: Svenska Vitterhetssamfundet, 138–44.
- 2004. "Monkey Business—or What is an Edition?" In: *Literary and Linguistic Computing* 19, 35–44.
- 2009. "... they hid their books underground". In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 113–25.
- Orme, Nicholas, ed. 1989. *Education and society in Medieval and Renaissance England*. London: Hambledon Press.



- Oxford English Dictionary Online*. 2012. Oxford: Oxford University Press. Available at: <http://www.oed.com/>. Accessed 7th May 2012.
- Page, Raymond I. 1992. "On the Feasibility of a Corpus of Anglo-Saxon Glosses: The View from the Library". In: *Anglo-Saxon Glossography: Papers read at the International Conference held in the Koninklijke Academie voor Wetenschappen Letteren en Schone Kunsten van België, Brussels, 8 and 9 September 1986*. Ed. by Rene Derolez. Brussels: Paleis der Academiën, 77–96.
- Pahta, Päivi. 1998. *Medieval Embryology in the Vernacular. The Case of De spermate*. Mémoires de la Société Néophilologique de Helsinki 53 53. (Diss.) Helsinki: Société Neophilologique.
- 2003. "On Structures of Code-Switching in Medical Texts from Medieval England". In: *Neuphilologische Mitteilungen* 104, 197–210.
  - 2004a. "Code-switching in medieval medical writing". In: *Medical and Scientific Writing in Late Medieval English*. Ed. by Irma Taavitsainen and Päivi Pahta. Cambridge: Cambridge University Press, 73–99.
  - 2004b. "So Seip Idem Comentator: Code-switching and Organization of Knowledge in John Trevisa's Translation of *De Proprietatibus Rerum*". In: *Voices on the Past: Studies in Old and Middle English Language and Literature*. Ed. by Alicia Rodríguez Álvarez and Francisco Alonso Almeida. A Coruña: Netbiblo, 35–48.
- Pahta, Päivi and Arja Nurmi. 2011. "Multilingual discourse in the domain of religion in medieval and early modern England: A corpus approach to research on historical code-switching". In: *Code-switching in Early English*. Ed. by Herbert Schendl and Laura Wright. Berlin: De Gruyter Mouton, 219–52.
- Pahta, Päivi and Irma Taavitsainen. 2004. "Vernacularisation of scientific and medical writing". In: *Medical and Scientific Writing in Late Medieval English*. Ed. by Irma Taavitsainen and Päivi Pahta. Cambridge: Cambridge University Press, 1–22.
- 2011. "An interdisciplinary approach to medical writing in Early Modern English". In: *Medical Writing in Early Modern English*. Ed. by Irma Taavitsainen and Päivi Pahta. Cambridge: Cambridge University Press, 1–8.
- Palmer, Carole L. 2004. "Thematic Research Collections". In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, Ray Siemens and John Unsworth. Malden, MA: Blackwell Publishing, 348–65.
- Parker, D. C. 2006. "Electronic religious texts: the Gospel of John". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 197–209.
- Parkes, M. B. 1969. *English Cursive Book Hands: 1250–1500*. Oxford: Clarendon Press.
- 1991a. "The influence of the concepts of *ordinatio* and *compilatio* on the development of the book". In: *Scribes, scripts and readers: studies in the communication, presentation and dissemination of medieval texts*. Ed. by M. B. Parkes. London: Hambledon Press, 35–70. Originally published in: Alexander, J. J. G. and M. T. Gibson (eds). 1976. *Medieval Learning and Literature: Essays Presented to R. W. Hunt*. Oxford: Oxford University Press.
  - 1991b. "The literacy of the laity". In: *Scribes, scripts and readers: studies in the communication, presentation and dissemination of medieval texts*. Ed. by M. B. Parkes. London: Hambledon Press, 275–98. Originally published in: Daiches D. and A. K. Thorlby (eds). 1973. *Literature and Western Civilization: The Medieval World*. London: Aldus Books.

- Parkes, M. B. 1992. *Pause and Effect: An Introduction to the History of Punctuation in the West*. Aldershot: Scolar Press.
- 2008. *Their Hands Before Our Eyes: A Closer Look at Scribes. The Lyell Lectures Delivered in the University of Oxford 1999*. Aldershot: Ashgate.
- Parkes, M. B. and A. I. Doyle. 1978. "The Production of Copies of the Canterbury Tales and the Confessio Amantis in the Early Fifteenth Century". In: *Medieval Scribes, Manuscripts and Libraries: Essays Presented to N. R. Ker*. Ed. by M. B. Parkes and A. G. Watson. London: Scolar Press, 163–210.
- Paston-Williams, Sara. 1993. *The Art of Dining: A History of Cooking & Eating*. London: The National Trust.
- Pawson, Dave, ed. 2012. *Extensible Stylesheet Language (XSL) Version 2.0*. W3C Working Draft. The World Wide Web Consortium. Available at: <http://www.w3.org/TR/2012/WD-xslfo20-20120117/>. Accessed 27th February 2012.
- Pearsall, Derek. 1985. "Editing Medieval Texts: Some Developments and Some Problems". In: *Textual Criticism and Literary Interpretation*. Ed. by Jerome J. McGann. Chicago: The University of Chicago Press, 92–106.
- 1989. Introduction. In: *Book Production and Publishing in Britain 1375-1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 1–10.
- 1992. "Authorial Revision in Some Late-Medieval English Texts". In: *Crux and Controversy in Middle English Textual Criticism*. Ed. by Alistair J. Minnis and Charlotte Brewer. Cambridge: D. S. Brewer, 39–48.
- 1994. "Theory and Practice in Middle English Editing". In: *TEXT: Transactions of the Society for Textual Scholarship* 7, 107–26.
- Pegge, Samuel, ed. 1780. *The Forme of Cury, a Roll of Ancient English Cookery. Compiled, about A. D. 1390, by the Master-Cooks of King Richard II, Presented afterwards to Queen Elizabeth, by Edward Lord Stafford, And now in the Possession of Gustavus Brander, Esq. Illustrated with Notes, And a copious Index, or Glossary*. London: J. Nichols.
- Piccard, Gerhard. 1961-1997. *Die Wasserzeichenkartei Piccard im Hauptstaatsarchiv Stuttgart: Findbuch*. 17 vols. Veröffentlichungen der Staatlichen Archivverwaltung Baden-Württemberg. Sonderreihe. Stuttgart: Kohlhammer. Available at: <http://www.piccard-online.de/>. Accessed 7th January 2014.
- Pichler, Alois. 1995a. "Advantages of a Machine-Readable Version of Wittgenstein's Nachlaß". In: *Culture and Value: Philosophy and the Cultural Sciences*. Ed. by K. Johannessen and T. Nordenstram. n.p.: The Austrian Ludwig Wittgenstein Society, 770–6.
- 1995b. "Transcriptions, Texts and Interpretation". In: *Culture and Value: Philosophy and the Cultural Sciences*. Ed. by K. Johannessen and T. Nordenstram. n.p.: The Austrian Ludwig Wittgenstein Society, 690–5.
- Pidd, Michael and Estelle Stubbs. 1997. "From Medieval Manuscripts to Electronic Text: A Transcriber's Tale". In: *The Canterbury Tales Project: Occasional Papers*. Ed. by Norman Blake and Peter Robinson. Oxford: Office for Humanities Communication, 55–9.
- Pierazzo, Elena. 2009. "Digital Genetic Editions: The Encoding of time in Manuscript Transcription". In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 169–86.
- 2011. "A rationale of digital documentary editions". In: *Literary and Linguistic Computing* 26, 463–77.

- Pierazzo, Elena and Peter A. Stokes. 2010. "Putting the Text back into Context: A Codicological Approach to Manuscript Transcription". In: *Codicology and Palaeography in the Digital Age 2*. Ed. by Franz Fischer, Christiane Fritze and Georg Vogeler. Norderstedt: Books on Demand, 397–429. Available at: [http://kups.ub.uni-koeln.de/4337/1/kpdz2online\\_gesamt.pdf](http://kups.ub.uni-koeln.de/4337/1/kpdz2online_gesamt.pdf). Accessed 29th December 2012.
- Pinkerton, John. 1797. *The history of Scotland from the accession of the House of Stuart to that of Mary. With appendixes of original papers*. Vol. 1. London: C. Dilly. Facsimile copy provided by *Eighteenth Century Collections Online (ECCO)*.
- Porter, Roy. 1992. "Introduction". In: *The Popularization of Medicine 1650–1850*. Ed. by Roy Porter. London: Routledge, 1–16.
- Portier, Pierre-Édouard and Sylvie Calabretto. 2009. "Methodology for the construction of multi-structured documents". In: *Proceedings of Balisage: The Markup Conference 2009*. Presented at Balisage: The Markup Conference 2009, Montréal, Canada, August 11 - 14, 2009. Available at: <http://www.balisage.net/Proceedings/vol3/html/Portier01/BalisageVol3-Portier01.html>. Accessed 9th May 2012.
- Power, Eileen, ed. 1992. *The Goodman of Paris (Le Ménagier de Paris): A Treatise on Moral and Domestic Economy by a Citizen of Paris c. 1393*. A reprint of the 1928 original. London: The Folio Society.
- Queneau, Raymond. 1961. *Cent mille milliards de poèmes*. Paris: Gallimard.
- Quirk, Randolph et al. 1991. *A Comprehensive Grammar of the English Language*. Ninth impression. London: Longman.
- Rawcliffe, Carole. 1995. *Medicine & Society in Later Medieval England*. Stroud: Sutton Publishing.
- Redon, Odile, Françoise Sabban and Silvano Serventi. 1998. *The Medieval Kitchen: Recipes from France and Italy*. Trans. by Edward Schneider. Chicago: University of Chicago Press.
- Rehbein, Malte. 2008. "The transition from classical to digital thinking: Reflections on Tim McLoughlin, James Barry and collaborative work". In: *Jahrbuch für Computerphilologie* 10. Ed. by Georg Braungart, Peter Gendolla and Fotis Janidis. Available at: <http://computerphilologie.de/jg08/rehbein.pdf>. Accessed 9th January 2013.
- Reimer, Stephen R. 2004. "Unbinding Lydgate's *Lives of Ss. Edmund and Fremund*". In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, 169–89.
- Renear, Allen. 2001. "Literal transcription—Can the text ontologist help?" In: *New Media and the Humanities: Research and Applications*. Ed. by Domenico Fiorimonte and Jon Usher. Oxford: Humanities Computing Unit, 23–30. Available at: [http://www.academia.edu/292930/New\\_Media\\_and\\_the\\_Humanities\\_Research\\_and\\_Applications](http://www.academia.edu/292930/New_Media_and_the_Humanities_Research_and_Applications). Accessed 4th January 2013.
- Renear, Allen H. 2004. "Text Encoding". In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, Ray Siemens and John Unsworth. Malden, MA: Blackwell Publishing, 218–39.
- Renfrow, Cindy. 1997. *Take a Thousand Eggs or More. A translation of medieval recipes from Harleian MS. 279, Harleian MS. 4016, and extracts of Ashmole MS. 1439, Laud MS. 553, and Douce MS. 55, with over 100 recipes adapted for modern cookery*. 2nd ed. 2 vols.

- Renn, Derek. 1991. *Kenilworth Castle*. Swindon: English Heritage.
- Rhys, John and J. Gwenogvryn Evans, eds. 1887. *The Text of the Mabinogion and Other Welsh Tales from the Red Book of Hergest*. Oxford: J. G. Evans.
- Richardson, Malcolm. 1980. "Henry V, the English Chancery, and Chancery English". In: *Speculum* 55, 726–50.
- Rissanen, Matti. 1989. "Three problems connected with the use of diachronic corpora". In: *ICAME Journal* 13, 16–9.
- 2000. "Standardisation and the language of early statutes". In: *The Development of Standard English 1300–1800: Theories, Descriptions, Conflicts*. Ed. by Laura Wright. Cambridge: Cambridge University Press, 117–30.
- 2008. "Corpus linguistics and historical linguistics". In: *Corpus Linguistics: An International Handbook*. Ed. by Anke Lüdeling and Merja Kytö. Berlin: Walter de Gruyter, 53–68.
- Robbins, Rossell Hope. 1970. "Medical Manuscripts in Middle English". In: *Speculum* 45, 57–73.
- Robins, William. 2004. "Toward a Disjunctive Philology". In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, 144–58.
- Robinson, Pamela R. 1980. "The 'Booklet': a self-contained unit in composite manuscripts". In: *Codicologica*, vol. 3: *Essais typologiques*. Ed. by Albert Gruys and J. P. Gumbert. Leiden: Brill, 46–69.
- Robinson, Peter. 1993. "Manuscript Politics". In: *The Politics of the Electronic Text*. Ed. by Warren Chernaik, Caroline Davis and Marilyn Deegan. Oxford: Office for Humanities Communication, Oxford University Computing Services, 9–16.
- 1994. "Collation, Textual Criticism, Publication, and the Computer". In: *TEXT: Transactions of the Society for Textual Scholarship* 7, 77–94.
- ed. 1996. *Chaucer: The Wife of Bath's Prologue on CD-ROM*. CD-ROM. Cambridge.
- ed. 2004. *The Miller's Tale on CD-ROM*. CD-ROM. Leicester.
- 2006. "The Canterbury Tales and other medieval texts". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 74–91.
- 2009. "What Text Really is Not, and Why Editors have to Learn to Swim". In: *Literary and Linguistic Computing* 24, 41–52.
- Robinson, Peter M. W. 1998. "The Computer and the Making of Editions". In: *A Guide to Editing Middle English*. Ed. by Vincent McCarren and Douglas Moffat. Ann Arbor: University of Michigan Press, 249–61.
- Robinson, Peter and Elizabeth Solopova. 1993. "Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue". In: *The Canterbury Tales Project: Occasional Papers*. Ed. by Norman Blake and Peter Robinson. Oxford: Office for Humanities Communication, 19–52.
- 2006. "Guidelines for Transcription of the Manuscripts of The Wife of Bath's Prologue". In: *The Nun's Priest's Tale on CD-ROM*. Ed. by Paul Thomas. Revised version of Robinson and Solopova (1993). Birmingham: Scholarly Digital Editions.
- Rodinson, Maxime. 1949. "Recherches sur les documents arabes relatifs à la cuisine". In: *Revue des Études Islamiques* XVII, 95–165.

- 1962. “La ma’muniyyat en orient et en occident”. In: *Études d’orientalisme dédiées à la mémoire de Lévi-Provençal*. Paris: Maisonneuve et Larose, 733–47.
- 2001a. “Ma’mūniyya East and West”. In: Rodinson, Maxime, A. J. Arberry and Charles Perry. *Medieval Arab Cookery*. Totnes: Prospect Books, 183–97.
- 2001b. “Studies in Arabic Manuscripts Relating to Cookery”. In: Rodinson, Maxime, A. J. Arberry and Charles Perry. *Medieval Arab Cookery*. Totnes: Prospect Books, 91–104.
- Rogos, Justyna. 2010. “Transcribing and editing graphetic detail in the manuscripts of Chaucer’s *Man of Law’s Tale*”. In: *Pe comoun peplis language*. Ed. by Marcin Krygier and Liliana Sikorska. Bern: Peter Lang, 79–86.
- Roland, Meg. 2004. “‘Als! Who may tryste thys world’: The Malory Documents and a Parallel-text Edition”. In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, 37–57.
- Romaine, Suzanne. 1982. *Socio-Historical Linguistics: Its Status and Methodology*. Cambridge: Cambridge University Press.
- 1995. *Bilingualism*. 2nd ed. Language in society 13. Oxford: Blackwell.
- Roos, Teemu and Tuomas Heikkilä. 2009. “Evaluating methods for computer-assisted stemmatology using artificial benchmark data sets”. In: *Literary and Linguistic Computing* 24, 417–33.
- Roos, Teemu, Tuomas Heikkilä and Petri Myllymäki. 2006. “A Compression-Based Method for Stemmatic Analysis. 17th European Conference on Artificial Intelligence, August 29–September 1, 2006, Riva Del Garda, Italy. Including Prestigious Applications of Intelligent Systems (PAIS 2006). Proceedings”. In: *ECAI 2006*. Ed. by Gerhard Brewka. IOS Press, 805–6. Available at: <http://www.cs.helsinki.fi/u/ttonteri/pub/ecai06.pdf>. Accessed 21st December 2013. An extended version of the published summary is available at.
- Rosenberg, Bob. 2006. “Documentary editing”. In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O’Brien O’Keeffe and John Unsworth. New York: The Modern Language Association of America, 92–104.
- Rothwell, William. 1985. “Stratford Atte Bowe and Paris”. In: *Modern Language Review* 80, 39–54.
- 1993. “The Legacy of Anglo-French: faux amis in French and English”. In: *Zeitschrift für romanische Philologie* 109, 16–46.
- 1998. “Arrivals and departures: The adoption of French terminology into middle English”. In: *English Studies* 79, 144–65.
- 2000. “Aspects of lexical and morphosyntactical mixing in the languages of medieval England”. In: *Multilingualism in Later Medieval Britain*. Ed. by David Trotter. Cambridge: D. S. Brewer, 213–32.
- 2004. “Henry of Lancaster and Geoffrey Chaucer: Anglo-French and Middle English in fourteenth-century England”. In: *Modern Language Review* 99, 313–27.
- 2007. “Synonymity and semantic variability in medieval French and Middle English”. In: *Modern Language Review* 102, 363–80.
- Roueché, Charlotte. 2009. “Digitizing Inscribed Texts”. In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 159–68.

- Ryan, W. F. and Charles B. Schmitt, eds. 1982. *The Secret of Secrets: sources and influences*. Warburg Institute Surveys 9. London: The Warburg Institute.
- Rymer, Thomas. 1729. *Fœdera, Conventiones, Literæ, Et Cujuscunque Generis Acta Publica, inter Reges Angliæ, Et Alios quosvis Imperatores, Reges, Pontifices, Principes, vel Communitates ab Ineunte Sæculo Duodecimo, viz. ab Anno 1101, ad nostra usque Tempora, Habita aut Tractata; Ex Autographis, infra Secretiores Archivorum Regiorum Thesaurarias, per multa Sæcula reconditis, fideliter Exscripta*. Vol. 9. London: J. Tonson. Facsimile copy provided by *Eighteenth Century Collections Online* (ECCO).
- Saenger, Paul. 1982. "Silent reading: its impact on late medieval script and society". In: *Viator* 13, 368–414.
- Samuels, M. L. 1972. *Linguistic Evolution, with Special Reference to English*. Cambridge: Cambridge University Press.
- 1989. "Some applications of Middle English dialectology". In: *Middle English Dialectology: Essays on Some Principles and Problems*. Ed. by Margaret Laing. Originally published in *English Studies* 44 (1963), 81–94. Aberdeen: Aberdeen University Press, 64–80.
- Sanders, Arnold. 1995. "Hypertext, Learning, and Memory: Some Implications from Manuscript Tradition". In: *TEXT: Transactions of the Society for Textual Scholarship* 8, 125–43.
- Sass, Lorna. 1975. *To the King's Taste: Richard II's Book of Feasts and Recipes Adapted for Modern Cooking*. New York: The Metropolitan Museum of Art.
- 1979. "A critical edition of The Fourme of cury: culinary vocabulary in late Middle English". PhD thesis. Columbia University.
- Schendl, Herbert. 2000. "Linguistic aspects of code-switching in medieval English texts". In: *Multilingualism in Later Medieval Britain*. Ed. by David Trotter. Cambridge: D. S. Brewer, 77–92.
- 2001. "Code-switching in medieval English poetry". In: *Language Contact in the History of English*. Ed. by Dieter Kastovsky and Arthur Mettinger. Frankfurt am Main: Peter Lang, 305–35.
- 2011. "Beyond boundaries: Code-switching in the leases of Oswald of Worcester". In: *Code-switching in Early English*. Ed. by Herbert Schendl and Laura Wright. Berlin: De Gruyter Mouton, 47–94.
- 2013. "Code-switching in medieval macaronic sermons". In: *Multilingualism in Medieval Britain: Sources and Analysis*. Ed. by Judith Jefferson and Ad Putter. Turnhout: Brepols, ??
- Schendl, Herbert and Laura Wright. 2011. "Code-switching in Early English: Historical background and methodological and theoretical issues". In: *Code-switching in Early English*. Ed. by Herbert Schendl and Laura Wright. Berlin: De Gruyter Mouton, 15–46.
- Schirmer, Walter F., ed. 1961. *John Lydgate: A study in the culture of the xv<sup>th</sup> century*. Trans. by Ann E. Keep. Berkeley: University of California Press.
- Schmidt, Desmond. 2009. "Merging Multi-Version Texts: a Generic Solution to the Overlap Problem". In: *Proceedings of Balisage: The Markup Conference 2009*. Presented at Balisage: The Markup Conference 2009, Montréal, Canada, August 11–14, 2009. Available at: <http://www.balisage.net/Proceedings/vol13/html/Schmidt01/BalisageVol13-Schmidt01.html>. Accessed 9th May 2012.

- Scott, Kathleen L. 1989. "Design, decoration and illustration". In: *Book Production and Publishing in Britain 1375-1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 31-64.
- Scully, Eleanor D. and Terence Scully. 1995. *Early French Cookery: Sources, History, Original Recipes and Modern Adaptations*. Ann Arbor: The University of Michigan Press.
- Scully, Terence, ed. 1988. *The Viandier of Taillevent: An Edition of all Extant Manuscripts*. Ottawa: University of Ottawa Press.
- 1992. *The Art of Cookery in the Middle Ages*. Woodbridge: Boydell & Brewer.
- ed. and trans. 1997. *The Vivendier. A Fifteenth-century French Cookery Manuscript*. Totnes: Prospect Books.
- Searle, Leroy F. 2004. "Emerging Questions: Text and Theory in Contemporary Criticism". In: *Voice, Text, Hypertext: Emerging Practices in Textual Studies*. Ed. by Raimonda Modiano, Leroy F. Searle and Peter Shillingsburg. Seattle: University of Washington Press, 3-21.
- Shaw, William Arthur and George Dames Burtchaell. 1906. *The knights of England; a complete record from the earliest time to the present day of the knights of all the orders of chivalry in England, Scotland, and Ireland, and of knights bachelors*. 2 vols. London: Sherratt and Hughes.
- Shillingsburg, Peter. 1986. *Scholarly Editing in the Computer Age: Theory and Practice*. Athens, Georgia: University of Georgia Press.
- 1991. "The Autonomous Author, the Sociology of Texts, and the Polemics of Textual Criticism". In: *Devils and Angels: Textual Editing and Literary Theory*. Ed. by Philip Cohen. Charlottesville: University Press of Virginia, 22-43.
- 1999. "Editing Determinate Material Texts". In: *TEXT: Transactions of the Society for Textual Scholarship* 12, 59-71.
- 2004. "Hagiolatry, Cultural Engineering, Monument Building, and Other Functions of Scholarly Editing". In: *Voice, Text, Hypertext: Emerging Practices in Textual Studies*. Ed. by Raimonda Modiano, Leroy F. Searle and Peter Shillingsburg. Seattle: University of Washington Press, 412-23.
- Shuffelton, George, ed. 2008. *Codex Ashmole 61: A Compilation of Popular Middle English Verse*. Kalamazoo: Medieval Institute Publication. Available at: <http://www.lib.rochester.edu/camelot/teams/sgas.htm>. Accessed 7th February 2012.
- SIL International. 2011. *Codes for the representation of names of languages - Part 3: Alpha-3 code for comprehensive coverage of languages*. ISO Standard 639-3. Dallas, TX: SIL International. Available at: <http://www.sil.org/iso639-3/default.asp>. Accessed 2nd January 2012.
- Simon, Eckehard. 2000. "Manuscript Production in Medieval Theatre: The German Carnival Plays". In: *New Directions in Later Medieval Manuscript Studies*. Ed. by Derek Pearsall. Rochester, NY: York Medieval Press, 143-65.
- Sinclair, John. 2004. "Language and Computing, Past and Present". In: *Proceedings of the 14<sup>th</sup> Europeans Symposium on Language for Special Purposes*. Ed. by Khurshid Ahmad and Margaret Rogers. Guildford: University of Surrey, 1-12.
- 2005. "Corpus and Text—Basic Principles". In: *Developing Linguistic Corpora: a Guide to Good Practice*. Ed. by Martin Wynne. Oxford: Oxbow Books, 1-16. Available at: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm>. Accessed 29th January 2014.

- Singer, Dorothea Waley and Annie Anderson. 1950. *Catalogue of Latin and Vernacular Plague Texts in Great Britain and Ireland in manuscripts written before the sixteenth century*. Collection de Travaux de l'Académie Internationale d'Histoire des Sciences 5. London: William Heinemann.
- Siraisi, Nancy G. 1990. *Medieval & Early Renaissance Medicine: An Introduction to Knowledge and Practice*. Chicago: University of Chicago Press.
- Skaarup, Bi. 1992. "Sources of medieval cuisine in Denmark". In: *Du Manuscrit à la table: Essais sur la cuisine au Moyen Âge et répertoire des manuscrits médiévaux contenant des recettes culinaires*. Ed. by Carole Lambert. Montreal: Les Presses de l'Université de Montreal, 39–43.
- Smith, Jeremy. 2008. "Issues of linguistic categorisation in the evolution of written Middle English". In: *Medieval Texts in Context*. Ed. by Graham D. Caie and Denis Renevey. London: Routledge, 211–24.
- Smith, Jeremy and Christian Kay. 2011. "The pragmatics of punctuation in Older Scots". In: *Communicating Early English Manuscripts*. Ed. by Päivi Pahta and Andreas H. Jucker. Cambridge: Cambridge University Press, 212–25.
- Smith, Martha Nell. 2004. "Electronic Scholarly Editing". In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, Ray Siemens and John Unsworth. Malden, MA: Blackwell Publishing, 306–22.
- Solopova, Elizabeth, ed. 2000. *Chaucer: The General Prologue on CD-ROM*. CD-ROM. Cambridge.
- Spencer, Judith, ed. and trans. 1984. *The Four Seasons of the House of Cerruti*. New York: Facts on File Publications.
- Sperberg-McQueen, C. M. 1991. "Text in the Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts". In: *Literary and Linguistic Computing* 6, 34–46.
- 2006. "Rabbit/Duck grammars: a validation method for overlapping structures". In: *Proceedings of Extreme Markup Languages, Montréal, Québec, 2006*. Available at: <http://conferences.idealliance.org/extreme/html/2006/SperbergMcQueen01/EML2006SperbergMcQueen01.html>. Accessed 9th May 2012.
  - 2007. "Representation of overlapping structures". In: *Proceedings of Extreme Markup Languages, Montréal, Québec, 2007*. Available at: <http://conferences.idealliance.org/extreme/html/2007/SperbergMcQueen01/EML2007SperbergMcQueen01.html>. Accessed 9th May 2012.
  - 2009. "How to teach your edition how to swim". In: *Literary and Linguistic Computing* 24, 27–39.
- Sperberg-McQueen, C. M. and Claus Huitfeldt. 2004. "GODDAG: A Data Structure for Overlapping Hierarchies. 8th International Conference on Digital Documents and Electronic Publishing, DDEP 2000, 5th International Workshop on the Principles of Digital Document Processing, PODDP 2000, Munich, Germany, September 13–15, 2000. Revised Papers". In: *Digital Documents: Systems and Principles*. Ed. by P. King and E. V. Munson. Springer.
- Sponsler, Claire. 2001. "Eating Lessons: Lydgate's 'Dietary' and Consumer Conduct". In: *Medieval Conduct*. Ed. by Kathleen Ashley and Robert L. A. Clark. Minneapolis: University of Minnesota Press, 1–22.
- Stannard, Jerry. 1982. "Rezeptliteratur as Fachliteratur. proceedings of the Special Session on Medieval Fachliteratur of the sixteenth International Congress on



- Medieval Studies, Kalamazoo, Michigan (U.S.A.), May 10, 1981". In: *Studies on medieval Fachliteratur*. Ed. by William Eamon. Brussel: UFSAL, 59–63.
- Steele, Robert, ed. 1898. *Three Prose Versions of the Secreta Secretorum*. Early English Text Society, Extra Series 74. London: K. Paul, Trench, Trübner & co.
- 1920. *Secretum Secretorum, cum glossis et notulis, tractatus brevis et utilis ad declarandum quedam obscure dicta Fratris Rogeri*. Oxford: Clarendon Press.
- Stenroos, Merja and Martti Mäkinen. 2009. *MEG-C Corpus Manual*. Version 2009.1. Stavanger: University of Stavanger. Available at: [http://www.uis.no/getfile.php/Forskning/Kultur/MEG/Corpus\\_manual\\_2009\\_1.pdf](http://www.uis.no/getfile.php/Forskning/Kultur/MEG/Corpus_manual_2009_1.pdf). Accessed 5th December 2012.
- Stevenson, Joseph, ed. 1837. *The life and death of King James the First of Scotland*. Glasgow: The Maitland Club.
- Stokes, Peter. 2010. "Teaching Manuscripts in the Digital Age". In: *Codicology and Palaeography in the Digital Age 2*. Ed. by Franz Fischer, Christiane Fritze and Georg Vogeler. Norderstedt: Books on Demand, 229–45. Available at: [http://kups.ub.uni-koeln.de/4337/1/kpdz2online\\_gesamt.pdf](http://kups.ub.uni-koeln.de/4337/1/kpdz2online_gesamt.pdf). Accessed 29th December 2012.
- 2011a. "Describing Handwriting, Part I". In: *DigiPal Blog*. Blog entry. Available at: <http://www.digipal.eu/blog/describing-handwriting-part-i/>. Accessed 25th June 2014.
- 2011b. "Describing Handwriting, Part II: Terminology". In: *DigiPal Blog*. Blog entry. Available at: <http://www.digipal.eu/blog/describing-handwriting-part-ii-terminology/>. Accessed 25th June 2014.
- 2011c. "Describing Handwriting, Part IV: Recapitulation and Formal Model". In: *DigiPal Blog*. Blog entry. Available at: <http://www.digipal.eu/blog/describing-handwriting-part-iv-recapitulation-and-formal-model/>. Accessed 25th June 2014.
- Stow, John. 1580. *The Chronicles of England, from Brute vnto this present yeare of Christ, 1580*. London: Henrie Bynneman. Facsimile copy provided by *Early English Books Online (EEBO)*.
- Street, Brian V. 1995. *Social Literacies: critical approaches to literacy in development, ethnography and education*. London: Longman.
- Strong, Roy. 2002. *Feast: A History of Grand Eating*. London: Jonathan Cape.
- Stubbs, William, ed. 1882. *Chronicles of the Reigns of Edward I and Edward II*. Vol. 1: *Rerum Britannicarum Medii Ævi Scriptores, or Chronicles and Memorials of Great Britain and Ireland during the Middle Ages*. 76. London: Longman & Co. Available at: <http://gallica.bnf.fr/ark:/12148/bpt6k50344f>. Accessed 13th August 2013.
- Sturges, Robert S. 1991. "Textual Scholarship: Ideologies of Literary Production". In: *Exemplaria* 3, 109–31.
- Stührenberg, Maik and Daniel Jettka. 2009. "A toolkit for multi-dimensional markup: The development of SGF to XStandoff". In: *Proceedings of Balisage: The Markup Conference 2009*. Presented at Balisage: The Markup Conference 2009, Montréal, Canada, August 11 - 14, 2009. Available at: <http://www.balisage.net/Proceedings/vol13/html/Stuhrenberg01/BalisageVol13-Stuhrenberg01.html>. Accessed 9th May 2012.
- Suarez, Michael. 2000. "In Dreams Begins Responsibility: Novels, Promises, and the Electronic Editor". In: *Textual Studies and the Common Reader: Essays in*

- Editing Novels and Novelists*. Ed. by Alexander Pettit. Athens, GA: University of Georgia Press, 160–79.
- Suhr, Carla. 2011. “Mémoires de la Société Néophilologique de Helsinki”. PhD thesis. Helsinki: University of Helsinki.
- Sutherland, Kathryn, ed. 2003. *Jane Austen’s Fiction Manuscripts: A Digital Edition*. Available at: <http://www.janeausten.ac.uk/>. Accessed 23rd May 2012.
- 2009. “Being Critical: Paper-based Editing and the Digital Environment”. In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 13–25.
- Swales, John H. 1990. *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Svenska Litteratursällskapet i Finland SLS, ed. 2008-. *Zacharias Topelius Skrifter*. Svenska Litteratursällskapet i Finland. Available at: <http://www.topelius.fi/index.php>. Accessed 21st May 2013.
- Symes, Carol. 2004. “The Boy and the Blind Man: A Medieval Play Script and Its Editors”. In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, 105–43.
- Taavitsainen, Irma. 2000. “Scientific language and spelling standardisation”. In: *The Development of Standard English 1300–1800: Theories, Descriptions, Conflicts*. Ed. by Laura Wright. Cambridge: Cambridge University Press, 131–54.
- 2001. “Middle English Recipes: Genre characteristics, text type features and underlying traditions of writing”. In: *Journal of Historical Pragmatics* 2, 85–113.
- 2004. “Scriptorial ‘house-styles’ and discourse communities”. In: *Medical and Scientific Writing in Late Medieval English*. Ed. by Irma Taavitsainen and Päivi Pahta. Cambridge: Cambridge University Press, 209–40.
- Taavitsainen, Irma and Päivi Pahta. 1995. “Scientific ‘Thought-Styles’ in Discourse Structure: Changing Patterns in a Historical Perspective”. In: *Organization in Discourse: Proceedings from the Turku Conference*. Ed. by Brita Wärvik, Sanna-Kaisa Tanskanen and Risto Hiltunen. Turku: University of Turku, 519–29.
- Talbot, C. H. and E. A. Hammond, eds. 1965. *The medical practitioners in medieval England: A biographical register*. London: Wellcome Historical Medical Library.
- Tanselle, G. Thomas. 1978. “The editing of historical documents”. In: *Studies in Bibliography* 31, 1–56.
- 2006. “Foreword”. In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O’Brien O’Keeffe and John Unsworth. New York: The Modern Language Association of America, 1–6.
- Taylor, Andrew. 1996. “Into his secret chamber: reading and privacy in late medieval England”. In: *The Practice and Representation of Reading in England*. Ed. by James Raven, Helen Small and Naomi Tadmor. Cambridge: Cambridge University Press, 41–61.
- 2004. “Editing Sung Objects: The Challenge of Digby 23”. In: *The Book Unbound: Editing and Reading Medieval Manuscripts and Texts*. Ed. by Siân Echard and Stephen Partridge. Toronto: University of Toronto Press, 78–104.
- Taylor, Ann et al. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. York. Electronic texts and manuals available from the Oxford

- Text Archive. Manuals and text information available online at <http://www-users.york.ac.uk/lang22/YCOE/YcoeHome.htm>.
- TEI Consortium. 2014. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Ed. by Lou Burnard and Syd Bauman. Version 2.6.0. Charlottesville, VA: Text Encoding Initiative Consortium. Available at: <http://www.tei-c.org/Vault/P5/2.6.0/doc/tei-p5-doc/en/Guidelines.pdf>. Accessed 21st July 2014.
- Tennison, Jeni. 2008. "Representing Overlap in XML". In: *Jeni's Musings*. Available at: <http://www.jenitennison.com/blog/node/97>. Accessed 10th May 2012.
- Tennison, Jeni and Wendell Piez. 2002. "Layered Markup and Annotation Language (LMNL)". In: *Proceedings of Extreme Markup Languages Conference, Montreal, Québec, 2002*. Available at: <http://conferences.idealliance.org/extreme/html/2002/Tennison02/EML2002Tennison02.html>. Accessed 9th May 2012.
- Terras, Melissa M. 2010. "Artefacts and Errors: Acknowledging Issues of Representation in the Digital Imaging of Ancient Texts". In: *Codicology and Palaeography in the Digital Age 2*. Ed. by Franz Fischer, Christiane Fritze and Georg Vogeler. Norderstedt: Books on Demand, 43–61. Available at: [http://kups.ub.uni-koeln.de/4337/1/kpdz2online\\_gesamt.pdf](http://kups.ub.uni-koeln.de/4337/1/kpdz2online_gesamt.pdf). Accessed 29th December 2012.
- Thaisen, Jacob. 2008. "The Trinity Gower D Scribe's Two Canterbury Tales Manuscripts Revisited". In: *Design and Distribution of Late Medieval Manuscripts in England*. Ed. by Margaret Connolly and Linne Mooney. York: York Medieval Press, 41–60.
- The Unicode Consortium. 2011. *The Unicode Standard*. Version 6.0.0. Mountain View CA: The Unicode Consortium. ISBN: 978-1-936213-01-6. Available at: <http://www.unicode.org/versions/Unicode6.0.0/>. Accessed 2nd January 2012.
- Thomas, Paul and Barbara Bordalejo, eds. 2006. *The Nun's Priest's Tale on CD-Rom*. CD-ROM. Leicester.
- Thompson, James Westfall. 1939. *The Medieval Library*. Chicago: The University of Chicago Press.
- Thoresby, Ralph. 1715. *Ducatus Leodiensis: or, the topography of the ancient and populous town and parish of Leedes, and parts adjacent in the West-Riding of the county of York. With The Pedegrees of many of the Nobility and Gentry, and other Matters relating to those Parts; Extracted from Records, Original Evidences, and Manuscripts. By Ralph Thoresby, F. R. S. To which is added, at the Request of several Learned Persons, A Catalogue of his Musaeum, with the Curiosities Natural and Artificial, and the Antiquities; particularly the Roman, British, Saxon, Danish, Norman, and Scotch Coins, with Modern Medals. Also A Catalogue of Manuscripts; the various Editions of the Bible, and of Books Published in the Infancy of the Art of Printing. With An Account of some unusual Accidents that have attended some Persons, attempted after the Method of Dr. Plot*. London: Maurice Atkins. Facsimile copy provided by *Eighteenth Century Collections Online* (ECCO).
- Toon, Thomas E. 1991. "Dry-Point Annotations in Early English Manuscripts: Understanding Texts and Establishing Contexts". In: *Annotation and Its Texts*. Ed. by Stephen A. Barney. New York: Oxford University Press, 74–93.

- Treharne, Elaine. 2011. "The vernaculars of medieval England, 1170–1350". In: *The Cambridge Companion to Medieval English Culture*. Ed. by Andrew Galloway. Cambridge: Cambridge University Press, 217–36.
- Trotter, David. 2000. "Introduction". In: *Multilingualism in Later Medieval Britain*. Ed. by David Trotter. Cambridge: D. S. Brewer, 1–6.
- 2003. "Oceano Vox: You never know where a ship comes from. On multilingualism and language-mixing in medieval Britain." In: *Aspects of Multilingualism in European Language History*. Ed. by Kurt Braunmüller and Gisella Ferraresi. Amsterdam: Benjamins, 15–33.
- 2010. "Bridging the Gap: The (Socio)linguistic Evidence of Some Medieval English Bridge Accounts". In: *The Anglo-Norman Language and its Contexts*. Ed. by Richard Ingham. York: York Medieval Press, 52–62.
- 2011. "Death, taxes and property: Some code-switching evidence from Dover, Southampton and York". In: *Code-switching in Early English*. Ed. by Herbert Schendl and Laura Wright. Berlin: De Gruyter Mouton, 155–89.
- Turville-Petre, Thorlac. 1987. "Editing *The Wars of Alexander*". In: *Manuscripts and Texts: Editorial Problems in Later Middle English Literature*. Ed. by Derek Pearsall. Cambridge: D. S. Brewer, 143–60.
- Tyrkkö, Jukka. 2011. "Fuzzy Coherence: Making sense of Continuity in Hypertext Narratives". PhD thesis. Helsinki: University of Helsinki. Available at: <http://urn.fi/URN:ISBN:978-952-10-7384-7>. Accessed 19th May 2013.
- Tyrkkö, Jukka, Ville Marttila and Carla Suhr. 2013. "The Culpeper Project: Digital editing of title pages". In: *Principles and Practices for the Digital Editing of Diachronic Data*. Ed. by Anneli Meurman-Solin and Jukka Tyrkkö. Helsinki: Research Unit for Variation, Contacts and Change in English. Available at: [http://www.helsinki.fi/varieng/series/volumes/14/tyrkkö\\_marttila\\_suhr/](http://www.helsinki.fi/varieng/series/volumes/14/tyrkkö_marttila_suhr/). Accessed 6th January 2014.
- UNESCO Education Sector. 2004. *The Plurality of Literacy and its implications for Policies and Programs*. UNESCO Education Sector Position Paper: 13. Paris: United Nations Educational, Scientific and Cultural Organization. Available at: <http://unesdoc.unesco.org/images/0013/001362/136246e.pdf>. Accessed 17th July 2013.
- Walsh, John A. 2010a. "Reaction to the Inadequacy of Embedded Markup". In: *Humanist* 23. Available at: <http://lists.digitalhumanities.org/pipermail/humanist/2010-April/001199.html>. Accessed 29th December 2012.
- Walsh, Marcus. 2010b. "Theories of Text, Editorial Theory, and Textual Criticism". In: *The Oxford Companion to the Book*. Ed. by Michael Suarez and H. R. Woudhuysen. Oxford: Oxford University Press, 156–63.
- Walther, Hans. 1963–69. *Proverbia sententiaeque Latinitatis Medii Aevi: lateinische Sprichwörter und Sentenzen des Mittelalters in alphabetischer Anordnung*. 6 vols. Carmina Medii Aevi posterioris Latina 2. Göttingen: Vandenhoeck & Ruprecht.
- Van Hulle, Dirk. 2006. "Authorial translation: Samuel Beckett's *Stirrings Still* / *Soubresauts*". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 150–60.
- Vander Meulen, David L. and G. Thomas Tanselle. 1999. "A System of Manuscript Transcription". In: *Studies in Bibliography* 52, 201–12.

- Vanhoutte, Edward. 1999. "Where Is the Editor? Resistance in the Creation of an Electronic Critical Edition". In: *Human IT* 3. Available at: <http://etjanst.hb.se/bhs/ith/1-99/ev.htm>. Accessed 5th June 2013.
- 2004. "An Introduction to the TEI and the TEI Consortium". In: *Literary and Linguistic Computing* 19, 9–16.
- 2006. "Prose fiction and modern manuscripts: limitations and possibilities of text encoding for electronic editions". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 161–80.
- 2009. "Every Reader his own Bibliographer". In: *Text Editing. Print and the Digital World*. Ed. by Marilyn Deegan and Kathryn Sutherland. Farnham: Ashgate, 99–110.
- Warner, Richard, ed. 1791. *Antiquitates Culinariæ, or Curious Tracts relating to the Culinary affairs of the Old English. With a preliminary discourse, Notes and Illustrations*. London: R. Blamire. Available at: <http://books.google.com/books?id=yew-AAAACAAJ>. Accessed 15th August 2013.
- Varvaro, Alberto. 1999. "The 'New Philology' from an Italian Perspective". In: *TEXT: Transactions of the Society for Textual Scholarship* 12, 49–58.
- Warwick, Claire. 2004. "Print Scholarship and Digital Resources". In: *A Companion to Digital Humanities*. Ed. by Susan Schreibman, Ray Siemens and John Unsworth. Malden, MA: Blackwell Publishing, 366–82.
- Watson, Andrew G., ed. 1966. *The Library of Sir Simonds D'Ewes*. London: The Trustees of the British Museum.
- Wedgwood, Josiah C. and Anne D. Holt. 1936. *History of Parliament: biographies of the members of the Commons House, 1439-1509*. London: H.M.S.O.
- Weinreich, Uriel, William Labov and Marvin I. Herzog. 1968. "Empirical Foundations for a Theory of Language Change". In: *Directions for Historical Linguistics: A Symposium*. Ed. by W. P. Lehmann and Yakov Malkiel. Austin, TX: University of Texas Press, 95–188.
- Weiss Adamson, Melitta. 1995. *Medieval Dietetics: Food and Drink in Regimen Sanitatis Literature from 800 to 1400*. Frankfurt am Main: Peter Lang.
- 2004. *Food in Medieval Times*. Westport, Connecticut: Greenwood Press.
- Weiss-Amer, Melitta. 1992. "The role of medieval physicians in the diffusion of culinary recipes and cooking practices". In: *Du Manuscrit à la table: Essais sur la cuisine au Moyen Âge et répertoire des manuscrits médiévaux contenant des recettes culinaires*. Ed. by Carole Lambert. Montreal: Les Presses de l'Université de Montreal, 69–80.
- Wenzel, Siegfried. 1986. "Reflections on (New) Philology". In: *Speculum* 61, 11–8.
- 1994. *Macaronic Sermons: Bilingualism and Preaching in Late-Medieval England*. Ann Arbor: Michigan University Press.
- Werlich, Egon. 1983. *A Text Grammar of English*. 2nd ed. Heidelberg: Quelle und Meyer.
- West, Martin L. 1998. "The Textual Criticism and Editing of Homer". In: *Editing texts: texte edieren*. Ed. by Glenn Most. Göttingen: Vandenhoeck & Ruprecht, 94–110.
- Wheatley, Henry B., ed. 1888. *Prices of Books: an inquiry into the changes in the price of books which have occurred in England at different periods*. The Library Series IV. London: George Allen.

- Wheaton, Barbara Ketcham. 1983. *Savoring the Past: the French kitchen and table from 1300 to 1789*. Philadelphia: University of Pennsylvania Press.
- White, Eileen. 2004a. "Domestic English Cookery and Cookery Books, 1575–1675". In: *The English Cookery Book: Historical Essays*. Ed. by Eileen White. Totnes: Prospect Books, 73–97.
- 2004b. "Preface". In: *The English Cookery Book: Historical Essays*. Ed. by Eileen White. Totnes: Prospect Books, 13–8.
- 2007a. "Preface". In: *The English Kitchen: Historical Essays*. Ed. by Eileen White. Totnes: Prospect Books, 9–10.
- 2007b. "Warm, Comforting Food: Soups, Broths and Pottages". In: *The English Kitchen: Historical Essays*. Ed. by Eileen White. Totnes: Prospect Books, 19–34.
- Wickersheimer, Ernest. 1936. *Dictionnaire biographique des médecins en France au Moyen Âge*. Vol. 1. Hautes études médiévales et modernes. Genève: Droz.
- Williams, Proctor William and Craig S. Abbott. 1999. *An Introduction to Bibliographical and Textual Studies*. New York: The Modern Language Association of America.
- Williams, Steven J. 1994. "Roger Bacon and His Edition of the Pseudo-Aristotelian *Secretum secretorum*". In: *Speculum* 69, 57–73.
- 2003. *The secret of secrets: the scholarly career of a pseudo-Aristotelian text in the Latin Middle Ages*. Ann Arbor: University of Michigan Press.
- Wilson, C. Anne. 2004. "An Introduction to the Cookery Book Collection in the Brotherton Library, University of Leeds". In: *The English Cookery Book: Historical Essays*. Ed. by Eileen White. Totnes: Prospect Books, 19–27.
- 2007. "Blancmange: A Tale of Seven Centuries". In: *The English Kitchen: Historical Essays*. Ed. by Eileen White. Totnes: Prospect Books, 43–56.
- Winkler, William E. 1990. "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage". In: *Proceedings of the Survey Research Methods Section, American Statistical Association*, 354–9. Available at: [http://www.amstat.org/sections/srms/Proceedings/papers/1990\\_056.pdf](http://www.amstat.org/sections/srms/Proceedings/papers/1990_056.pdf). Accessed 8th October 2013.
- Visser, Margaret. 1991. *The Rituals of Dinner: The Origins, Evolution, Eccentricities, and Meaning of Table Manners*. New York: Grove Weidenfeld.
- Witern, Christian. 2006. "Writing systems and character representation". In: *Electronic Textual Editing*. Ed. by Lou Burnard, Katherine O'Brien O'Keeffe and John Unsworth. New York: The Modern Language Association of America, 291–8.
- Wittgenstein, Ludwig. 2000. *Wittgenstein's Nachlass: Text and Facsimile Version. The Bergen Electronic Edition*. Published on CD-ROM. Oxford: Oxford University Press.
- Voigts, Linda Ehrsam. 1984. "Medical Prose". In: *Middle English Prose: a critical guide to major authors and genres*. Ed. by A. S. G. Edwards. New Brunswick, NJ: Rutgers University Press, 315–35.
- 1989. "Scientific and medical books". In: *Book Production and Publishing in Britain 1375–1475*. Ed. by Jeremy Griffiths and Derek Pearsall. Cambridge: Cambridge University Press, 345–402.
- 1990. "The 'Sloane group': Related Scientific and Medical Manuscripts from the Fifteenth Century in the Sloane Collection". In: *The British Library Journal* 16, 26–57.

- 1996. "What's the Word? Bilingualism in Late-Medieval England". In: *Speculum* 71, 813–26.
- Wood, Margaret. 1983. *The English Mediaeval House*. London: Bracken Books.
- Woolgar, Christopher M., Dale Serjeantson and Tony Waldron, eds. 2006. *Food in Medieval England: Diet and Nutrition*. Medieval History and Archaeology. Oxford: Oxford University Press.
- Wright, C. E. and Ruth C. Wright, eds. 1966. *The Diary of Humfrey Wanley, 1715–1726*. 2 vols. London: The Bibliographical Society.
- Wright, Laura. 1992. "Macaronic writing in a London archive". In: *History of Englishes. New Methods and Interpretations in Historical Linguistics*. Ed. by Matti Rissanen et al. Berlin: Mouton de Gruyter, 762–70.
- 1995. "A hypothesis on the structure of macaronic business writing". In: *Medieval Dialectology*. Ed. by Jacek Fisiak. Berlin: Mouton de Gruyter, 309–21.
- 1996. "About the evolution of Standard English". In: *Studies in English Language and Literature: 'Doubt wisely': Papers in honour of E. G. Stanley*. Ed. by M. J. Toswell and E. M. Tyler. London: Routledge, 99–115.
- 1999. "Mixed-language business writing: five hundred years of code-switching". In: *Language Change: Advances in Historical Sociolinguistics*. Ed. by Ernst Håkon Jahr. Berlin: Mouton de Gruyter, 99–117.
- 2000a. "Bills, accounts, inventories: everyday trilingual activities in the business world of later medieval England". In: *Multilingualism in Later Medieval Britain*. Ed. by David Trotter. Cambridge: D. S. Brewer, 149–56.
- 2000b. "Introduction". In: *The Development of Standard English 1300–1800: Theories, Descriptions, Conflicts*. Ed. by Laura Wright. Cambridge: Cambridge University Press, 1–8.
- 2010. "A Pilot Study on the Singular Definite Articles *le* and *la* in Fifteenth-Century London Mixed-Language Business Writing". In: *The Anglo-Norman Language and its Contexts*. Ed. by Richard Ingham. York: York Medieval Press, 130–42.
- 2011. "On variation in medieval mixed-language business writing". In: *Code-switching in Early English*. Ed. by Herbert Schendl and Laura Wright. Berlin: De Gruyter Mouton, 191–218.
- Wyatt, Diana. 1987. "Editing for REED". In: *Manuscripts and Texts: Editorial Problems in Later Middle English Literature*. Ed. by Derek Pearsall. Cambridge: D. S. Brewer, 161–70.
- Wynne, Martin. 2005. "Archiving, distribution and preservation". In: *Developing Linguistic Corpora: a Guide to Good Practice*. Ed. by Martin Wynne. Oxford: Oxbow Books, 71–8. Available at: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter6.htm>. Accessed 29th January 2014.
- Yergeau, François et al., eds. 2008. *Extensible Markup Language (XML) 1.0*. 5th ed. W3C Recommendation. The World Wide Web Consortium. Available at: <http://www.w3.org/TR/2008/REC-xml-20081126/>. Accessed 2nd January 2012.
- Zeller, Hans. 1975. "A New Approach to the Critical Constitution of Literary Texts". In: *Studies in Bibliography* 28, 231–64.
- 1995a. "Record and Interpretation: Analysis and Documentation as Goal and Method of Editing". In: *Contemporary German Editorial Theory*. Ed. by Hans Walter Gabler, George Bornstein and Gillian Borland Pierce. Ann Arbor: University of Michigan Press, 17–58.

- Zeller, Hans. 1995b. "Structure and Genesis in Editing: On German and Anglo-American Textual Criticism". In: *Contemporary German Editorial Theory*. Ed. by Hans Walter Gabler, George Bornstein and Gillian Borland Pierce. Ann Arbor: University of Michigan Press, 95–123.
- Öhrnberg, Kaj and Sahban Mroueh, eds. 1987. *Ibn Sayyār al-Warrāq: Kitāb al-ṭabīkh*. Helsinki: Societas Orientalis Fennica.



# Appendices



# Appendix A

## TEI XML Source Documents

As described in chapter 11, the edition of the six recipe collections making up the *Potage Dyvers* family as a data archive is encoded as a combination of six *base data files*, one for each manuscript version, along with a number of *annotation overlay files*, linked to one or more base of the data files. These files, which are found in the folder /Appendix A - Edition Source on the CD-ROM accompanying the printed edition of this thesis or in the Digital\_appendices.zip file accompanying the digital edition, include the following:<sup>1</sup>

**MS\_identifier.xml** the base data file for the *PD* manuscript version indicated by the MS\_identifier (see chapter 11)

**MS\_identifier\_expansion.xml** an editorial annotation overlay that contains editorial expansions of all abbreviated words in the *PD* manuscript version indicated by the MS\_identifier (see subsection 11.7.2)

**MS\_identifier\_normalisation+wordclass.xml** a non-editorial linguistic annotation overlay that contains normalised forms and word class information for all words in the *PD* manuscript version indicated by the MS\_identifier (see subsection 11.9.1)

**PD\_explanatory\_notes.xml** an editorial annotation overlay that contains explanatory notes linked to entire recipes or individual words or phrases in one or more of the *PD* manuscript versions (see subsection 11.8.2)

**PD\_recipe\_versions.xml** a non-editorial intertextual annotation overlay that links together *parallel versions* (as defined in subsection 10.3.2) of each unique recipe occurring in one or more of the *PD* manuscript versions (see subsection 11.9.2)

**PD\_canonical\_titles.xml** a non-editorial intertextual annotation overlay that associates each unique recipe occurring in one or more of the *PD* manuscript versions (see subsection 11.9.2) with a canonical descriptive title based on Hieatt and Nutter (2006) (see subsection 11.9.4)

**PD\_recipe\_sources.xml** a non-editorial intertextual annotation overlay that groups the 271 unique recipes occurring in one or more of the *PD* manuscript versions (see subsection 11.9.2) into six groups based on their likely textual

---

<sup>1</sup> The placeholder MS\_identifier is used to represent the identifiers of the six manuscript versions (i.e. *MSAdditional5467*, *MSAshmole1439*, *MSCosinViii11A*, *MSDouce55*, *MSHarley279* and *MSHarley4016*) in the names of those files of which there is one for each of the six manuscript versions.

sources, discovered through an analysis (see chapter 13) of their occurrence patterns in the six manuscripts (see subsection 11.9.3)

## A.1 ODD and schema

In addition to the detailed prose documentation of the specific subset of TEI XML annotation used in the base data files and annotation overlays in chapter 11, it is also documented formally by a TEI One Document Does it All (ODD) file. The ODD file, whose format and use is defined in the *TEI Guidelines* (TEI Consortium 2014: 643–700), defines the specific components and elements of the TEI specification that are used in the documents making up the edition (along with the attributes and attribute values allowed for them), as well as the ways in which the annotation used in the file differs from the TEI specification. In addition to the TEI ODD files, which should be considered the normative documentation for the edition (consisting of the base data files and the different types of annotation overlays) the CD-ROM also contains a Relax NG Schema file, generated from the ODD file using the TEI Roma tool (<<http://www.tei-c.org/Roma/>>), which has been used to validate the source documents of this edition and can also be used to verify that any extensions made to the edition (new manuscript texts or annotation overlays) are formally correct.

The One Document Does it All (ODD) and Schema files for the edition, named PD\_ODD.xml and PD\_Schema.rng, are found in the folder /Appendix A - Edition Source/Schema.

## Appendix B

# Diplomatic transcriptions

The first of the three editorial outputs included in the present edition consists of a diplomatic representation of each manuscript page of all the six *PD* versions, reproducing not only the textual content of the original document page as a graphemic transcription, but also its physical layout and all other visual and physical features annotated into the base data file and described in sections 11.4, 11.5 and 11.6. The diplomatic representation of the text has been implemented by generating presentation-oriented XHTML files of each manuscript page through a series of XSLT transformations, formatting them using CSS and printing them into a single PDF file, found in the /Appendix B - Diplomatic Transcriptions folder of the accompanying CD-ROM, along with the XML source file from which the appendix is typeset.<sup>1</sup> The XSLT transformations used to generate the XHTML files (and the CSS files used to visually format them) are also included on the CD-ROM and described in appendix E.

### B.1 Presentation conventions

The textual content of the document is presented in its entirety, including all later additions and annotations, as well as any *forme work*, whether original or later. All words are presented in their abbreviated form and word-separation is not modernised but reflects that of the original, although in a binary form. All abbreviation markers and other special symbols are represented using a specially created font.<sup>2</sup> The layout of the original document page, annotated into the base data file using the system described in section 11.3 is represented by positioning the various textual elements of the page similarly to the original document. The size and proportions of the text block and the margins approximate those of the original document, but are not reproduced to scale, and the positions of the textual and graphical elements on the page do not replicate the absolute measurements of the

<sup>1</sup> The XHTML files were printed into individual PDF files using *wkpdf* (<<http://plessl.github.io/wkpdf/>>), combined into a single PDF file and cropped to remove superfluous margins using *PDF Scissors* (<<http://www.pdfscissors.com/>>).

<sup>2</sup> This font, based on the Junicode font by Peter S. Baker (<<http://junicode.sourceforge.net/>>) distributed under the Open Font Licence, is included with the browsable HTML edition (appendix D), similarly under the Open Font License.

original but are relative to each other and the different regions of the page. The visual representations used for the other textual and visual features of the document are described below under separate subsections, organised similarly to the description of their annotations in chapter 11.

### B.1.1 Scribal hands, emendations and annotations

The different scribal hands used in the manuscript are indicated visually by the use of colour, the colour used for the text corresponding roughly to the colour of the ink or pigment in the original.<sup>3</sup> The principal hand of the original document as well as the hands used for regular features such as *forme work* are described in the TEI header of the source file (see subsections 11.1.6 and 11.4.1) and in section 9.2, and are not described further in the transcription. All emendations (additions, deletions and substitutions) and annotations made in hands other than the original scribal one are indicated as such by a footnote generated procedurally based either on the annotation of the emendation or on a separate textual note (in complex cases) and keyed to the end of the appropriate text segment, using a descriptive phrase—also occurring at the beginning of the hand’s description—to identify each hand used in the document.

All emendations made in the text are indicated graphically at the position where they occur. Marked deletions (see subsection 11.4.2) are represented by the appropriate marking (underlining, strikethrough, subpunction, etc.) in the colour representing the hand in which the deletion is considered to have occurred, and are also described by a generated footnote. Additions made to the text are indicated by placing the added text to the appropriate position, together with any symbols used to indicate the point of insertion, and also described by a generated footnote indicating the hand in which the addition was made. Substitutions involving implicit deletions or additions are also described by a footnote, with any markings involved being also represented graphically. Substitutions involving overwriting are indicated by placing the added text on top of the deletion, while erased deletions—whether replaced by overwritten text or not—are indicated by printing the erased text (if still discernible in the original) with 10% opacity, both being also described by a footnote.

All annotations made to the text, as well as any *forme work*, are placed at the position in which they occur in the original document. For annotations made in a hand different from the surrounding text (usually the original scribal one), the hand is also indicated by a procedurally generated footnote.

### B.1.2 Highlighting and decoration

All visual highlighting and decoration found in the original document is also visually represented in the transcription. Super- and subscripted characters are displayed as such regardless of their function (whether indicating an abbreviation or not), and letters whose size does not correspond with their shape (i.e. ‘small capitals’ and ‘large minuscules’, described in subsection 11.5.1) are indicated using a

<sup>3</sup> The use of colour in this editorial output is one reason why it cannot be conveniently represented in printed form at reasonable cost and is only included in digital form, the other being the large number of pages required by it.

larger or smaller font size, as appropriate. Characters with decorative ascenders or descenders, mainly occurring on the top and bottom lines of the page, are indicated using a special version of the font with elongated ascenders or descenders.<sup>4</sup> Text that has been written in a coloured pigment (red or blue in the present edition) are represented using that colour, and characters that have been highlighted by touching them up with a stroke of coloured pigment are represented by a ‘drop shadow’ effect in that colour.

Text that has been underlined in the original or surrounded by a ‘frame’ on one or more sides is represented graphically by lines—the colour of which approximate the colour of the ink or pigment used—surrounding the highlighted segment on the appropriate sides. Drop capitals and large capitals with a height of multiple lines are represented by characters of the appropriate size, large capitals standing on the line and drop capitals being aligned to the top of the line (unless they have been offset in the original, in which case they have also been offset in the transcription). Similarly, segments of text written in a clearly different size in the original are represented by text visually approximating their size in relation to the surrounding text. Characters written in a script differing from that normally used by the hand in question, which in the present edition include the Uncial-style or ‘Lombardic’ capitals, are represented using a different, Uncial-style font.<sup>5</sup> Decorative ornaments such as line fillers or simple shapes drawn in empty parts of the page have been represented visually using SVG, while more complex drawings have been described in the textual notes, presented here as footnotes.

### B.1.3 Damage and legibility

Physical damage to the manuscript page is both represented visually and described by procedurally created footnotes. Damage affecting the textual content of the page is visually indicated by a gray background colour, the intensity of which indicates the severity of the damage, ranging from very light to dark gray. In addition to the visual representation, all damaged sections are also described by a footnote indicating the source and severity of the damage. Instances of damage that do not cover any text but are nevertheless annotated as relevant, as described in subsection 11.6.1, are indicated by footnotes attached to the place where the damage occurs. Passages that have been rendered unclear due to damage are described in the footnote describing the damage itself, while ones that are unclear for some other reason are described by a separate footnote.

Gaps in the text, whether caused by damage or deletion, are indicated in several slightly different ways depending on their extent and reason. Gaps caused by damage are indicated by empty space with a black background; if the extent of the gap is less than a single line, the width of the black area represents the number of characters that are estimated to be missing, and if the gap covers multiple lines,

<sup>4</sup> The shape of these decorated ascenders and descenders is not intended to represent the exact shape of the original document but merely to indicate the presence of a visually distinctive feature. Like the main font used for the diplomatic representation of the text in this editorial rendition and in the browsable HTML edition, these special versions of the font (UnicodeDECL\_AD-Regular.ttf and UnicodeDECL\_DD-Regular.ttf) are included with the browsable HTML edition (appendix D).

<sup>5</sup> As with the characters decorated by enlarged ascenders or descenders, this font is not intended to represent the exact shapes of the original, merely the fact that the style of the characters differs from the surrounding text.

the number of missing lines is indicated within square brackets in the middle of the black area representing the gap. Entire folia that have been lost due to damage are indicated by a footnote at the end of the preceding page (the verso of the previous folio) and the head of the following page. Gaps caused by deletion are indicated similarly, but with the empty space having a white background and being surrounded by a black border. Gaps left in the transcription due to illegibility of the text are indicated by replacing unclear characters by a question mark (?).



## Appendix C

# Parallel reading edition

One of the motivations for preparing an edition with full transcriptions of several versions of a single text is the ability to compare those versions to each other. For this purpose, the relationships between the recipes of the different manuscript versions have been established and annotated as described under subsection 11.9.2, which allows, among other things, the parallel presentation of the different versions of each unique recipe in the *Potage Dyvers* family. The second of the three editorial outputs included in the present edition consists of a parallel reading edition of all the 371 unique recipes contained in the six collections of the *Potage Dyvers* family. This parallel edition presents all of the parallel versions of each unique recipe next to each other, allowing the interpretation of semantically unclear or ambiguous passages in one of the versions to be supported by the other versions which may provide clearer readings, as well as the convenient comparison of their textual content. In order to facilitate the reading of the recipes for their content, this edition omits all visual features of the original document—except for line and page breaks which are indicated in the text—expands all abbreviations and implements all corrections made to the text, whether by the original scribe or later correctors. Explanatory notes commenting on various aspects of the recipes, including differences between the parallel versions, textual cruxes and obscure passages, are presented after each recipe, keyed to the appropriate points in the different versions. The parallel reading edition is included as a PDF document (Parallel\_edition.pdf) found in the /Appendix C - Parallel Reading Edition folder on the enclosed CD-ROM, along with the XML source file it is typeset from. The XSLT transformations used to generate the XML source files of the parallel reading edition and its subsequent  $\text{\LaTeX}$  rendition are also included on the CD-ROM and described in appendix E.

### C.1 Presentation conventions

Since the recipes occur in anything from one to all six of the manuscript versions and the order of the recipes varies considerably between the different manuscript versions, their parallel presentation necessitates their removal from the original manuscript context. This means that all content in the manuscripts that is not

directly connected to the recipes (i.e. folio numbers, catchwords, tables of contents, bills of fare, etc.) has been omitted from this parallel reading edition and the recipes presented in an editorially determined order, described below under subsection C.1.1.<sup>1</sup> In accordance with the purpose of this parallel presentation, the text of the recipes is formatted as what could be described as a *conservative reading edition*, which seeks to make it easily readable while preserving its linguistic characteristics. The treatment of the various textual (and visual) features found in the original manuscripts and annotated in the digital edition is described under subsection C.1.2 below.

### C.1.1 Order of presentation

Since both the range of recipes included and their ordering varies significantly between the different manuscript versions, there is no single ‘natural’ order for presenting all of the unique recipes contained in them. In the absence of an obvious order for all the recipes, this parallel reading edition organises the recipes on the basis of the structural analysis of the six MS versions reported in chapter 13, dividing them into six groups based on their apparent sources. The  $\alpha$  group, presented first, contains all of the recipes that seem to have originally occurred in all six of the versions and most likely derive from a shared ancestor, forming the core of the *Potage Dyvers* family. The  $b$  group, presented next, consists of the recipes shared exclusively by the two closely related pairs of MSS, i.e. Ad, D, C and H4016, apart from seven recipes which also occur in MS As. The  $a$  group, which is the largest of the groups, consists of the recipes that are exclusive to MSS As and H279 and can thus be considered to be more peripheral to the *PD* family. Groups  $c$  and  $d$  contain even more peripheral recipes, the former being exclusive to MS As and the latter to MS H4016. The final group consists of those individual recipes whose textual origins are indeterminate. Within these groups, the ordering of the recipes follows the MS version containing the largest number of recipes in the group in question, which is MS D for groups  $\alpha$  and  $b$ , and MS H279 for group  $a$ .<sup>2</sup> For comparing the order of the recipes in the different MS versions, the reader is referred to section C.2 which provides six tables, each listing the order of the recipes in one of the manuscript versions (with additional information), and to the interactive table of contents of the browsable parallel reading edition (appendix D).

### C.1.2 Treatment of textual features

Since the purpose of the parallel reading edition is to present all the parallel versions in a readable yet linguistically faithful form, it omits many aspects of the

<sup>1</sup> The entire contents of each manuscript, represented as a diplomatic transcription, can be found in appendix B. Although most of the bills of fare contained in the different manuscript versions have been previously published in Austin (1888), Hieatt and Butler (1985) and Hieatt (2004), all of them are also reproduced in this thesis both in an easily readable format in section C.3 and as a part of the diplomatic transcript of each MS version.

<sup>2</sup> In the browsable parallel reading edition included in appendix D, the reader can follow the organisation of any of the manuscript versions, and the dynamic user interface planned for the eventual online version of this edition will provide even more organisational options.

original manuscript text. Of the original document layout, the only features represented are line and page breaks, the former marked by a double vertical line (||) and the latter by the folio number in square brackets (e.g. [f.4r]). The titles of the recipes are placed at the head of the recipe text and printed in bold face, regardless of their original location and formatting. Any labels added to the recipes, such as recipe numbers added by later annotators or preliminary labels added by the scribe as an instruction to the rubricator and not subsequently erased, are inserted before the title of the recipe (if it exists) and printed in lighter type. Any other annotations added by later users of the manuscript are omitted, as well as any headings or other material that is not connected with an individual recipe.

Physical aspects of the original manuscript like damage and visual highlighting or decoration are not represented in this reading edition. Damage is taken into account only in cases where it has resulted in the loss of text. Missing sequences of letters or other characters are indicated by a series of asterisks enclosed within square brackets (e.g. [\*\*\*\*\*]), while entire missing lines or parts of recipes are indicated by an italicized note within square brackets. Emendations (additions, deletions and substitutions) to the text itself are incorporated into the text, regardless of whether they were made in the original scribal hand or the hand of some later corrector. Deletions are simply omitted,<sup>3</sup> while additions are indicated by a combination of square brackets and slashes, using the annotation [/ . . \] for additions made on the line (whether added to an empty space or written over deleted text) and [\ . . /] for interlineal and marginal additions.<sup>4</sup>

In terms of the representation of the text itself, the editorial practices described in section 10.2 are observed. In brief, this means that the text is normalized to the graphemic level, word spacing is preserved but normalized to a binary form (ambiguous cases being weighted towards the norm), and the original manuscript spelling of words is preserved. For the reading edition, abbreviations have been expanded and indicated by italics.<sup>5</sup> Apart from abbreviation markers, all characters not included in the modern alphabet—including punctuation symbols, the letters *p* and *z* and astrological symbols are included using their Unicode or MUFI representations (see subsection 11.5.3), and non-abbreviating super- and subscript characters are represented as such.

### C.1.3 Editorial notes

Like the manuscript features described above, the textual notes describing the features of the original manuscript documents and texts are not included in the reading edition, but rather in the diplomatic transcriptions (appendix B). However, each group of parallel versions is accompanied by a selection of explanatory notes by the editor, which are conversely not included in the diplomatic transcriptions. Notes referring to specific passages in the recipes are keyed to the relevant passages in the recipes by superscript numbers inserted at the end of the passage and

<sup>3</sup> This treatment means that the deletion of longer spans extending over multiple lines results in the presence of several consecutive line breaks in the presentation.

<sup>4</sup> This annotation does not distinguish between additions made in the original scribal hand and in the hands of later correctors; for this information, the reader is referred to the diplomatic transcriptions of the manuscript versions in appendix B.

<sup>5</sup> Abbreviation markers which have been judged as otiose have simply been omitted.

at the head of the corresponding note. For notes that relate to several of the parallel versions, the same marker is inserted to the appropriate place on each of the relevant versions. Notes that pertain to the recipe as a whole are included at the head of the list of notes and are not preceded by a reference number.

## C.2 Tables of recipes in each of the manuscripts

To compensate for the editorial ordering of the recipes in the parallel reading edition, it is followed by a series of six tables, each of which presents the recipes in the order they occur in one of the manuscript versions. Each table contains the following columns:

- 1) the ordinal number of the recipe in the manuscript,
- 2) the title of the recipe in the manuscript (if one exists),
- 3) the location of the recipe in the manuscript,
- 4) the recipe reference, shared between all parallel versions of the same recipe,
- 5) the ‘source group’ of the recipe, used as the primary organising principle of the parallel reading edition, and
- 6) a list of the parallel versions found in each of the other manuscripts (complete with the ordinal number of the recipe in that manuscript in parentheses).

The following abbreviations are used for the different manuscripts:

<b>Ad</b>	London, British Library MS Additional 5467
<b>As</b>	Oxford, Bodleian Library MS Ashmole 1439
<b>C</b>	Durham, University Library MS Cosin V.iii.11
<b>D</b>	Oxford, Bodleian Library MS Douce 55
<b>H279</b>	London, British Library MS Harley 279
<b>H4016</b>	London, British Library MS Harley 4016

## C.3 Bills of fare included with the *Potage Dyvers*

The parallel reading edition of the recipes themselves is also supplemented with a reading edition of the bills of fare that have been appended to five of the six manuscript versions of the *Potage Dyvers* family. Similarly to the parallel reading edition itself, the bills of fare are here formatted as a *conservative reading edition* which seeks to make them readable while preserving their linguistic characteristics. Due to the formal list-like nature of the bills of fare, their original document layout is not represented here but replaced by a standardised list format.<sup>6</sup> The headings of each bill of fare and of their individual courses are printed in bold. As with the parallel reading edition, that physical aspects of the original manuscript such as damage and visual highlighting or decoration are not represented in this reading edition. Emendations (additions, deletions and substitutions) to the text itself are incorporated into the text, regardless of whether they were made in the original scribal hand or the hand of some later corrector. Deletions are simply

<sup>6</sup> For a representation of their layout and other visual properties, see the diplomatic transcriptions of the individual manuscript versions in appendix B.

omitted, while additions are indicated by a combination of square brackets and slashes, using the annotation [/ . . . \] for inline additions and [\ . . . /] for interlineal and marginal additions, similarly to the parallel reading edition described in appendix C.

As with the parallel reading edition, the textual content of the bills of fare is treated as described in section 10.2, i.e. the text is normalized to the graphemic level, word spacing is preserved but normalized to a binary form (ambiguous cases being weighted towards the norm), and the original manuscript spelling of words is preserved. As with the reading edition, abbreviations have been expanded and indicated by italics, and characters not included in the modern alphabet are included using their Unicode or MUFI representations.



## Appendix D

# Browsable HTML edition

The third editorial output included in this edition is the one that is properly native to the digital medium and could be considered a precursor to a proper online interface for the edition, even though it does not yet feature any kind of search functionality and also the navigational apparatus is extremely simplistic. This editorial output combines the previous two outputs by linking the parallel presentation of each unique recipe to those pages of the diplomatic transcriptions that contain the relevant recipe in each manuscript, and linking each recipe in the diplomatic transcription to the parallel presentation of all versions of that recipe.

In addition to these links joining the two representations, the diplomatic transcriptions contain links to the preceding and following page, as well as a hyper-linked index of all the pages allowing direct access to any page. The parallel presentation of each recipe is similarly linked to the parallel presentations of the preceding and following recipe in *each* of the versions in which it occurs, allowing the user to follow the order of recipes in any of the versions.<sup>1</sup> In addition to the sequential links, the HTML version of the parallel reading edition also contains a customisable table of contents which aligns all the parallel versions of the recipes horizontally and can be reorganised to reflect the order of recipes in any of the six manuscripts. In order to facilitate access to the metadata describing the manuscript versions, links are also provided to HTML renditions of the manuscript descriptions included in the metadata header of the base data files.

Unlike the two preceding outputs which are presented in essentially printed format as PDF files, this third editorial output is presented as a collection of inter-linked HTML files, viewable using any modern browser. These files are contained in the /Appendix D - HTML Edition/files folder, with a main table of contents serving as the navigational starting point for the edition being found in the /Appendix D - HTML Edition folder itself. While this static HTML representation does not offer any search functionality or allow the dynamic reformatting of the text according to user choices, it does provide an example of one possible navigational structure that in itself is already more useful than any printed representation. First of all, it avoids the problem of having to impose an arbitrary linear order on the recipes by presenting them as a multiply organised rhizomatic network, and sec-

---

<sup>1</sup> The currently 'active' version, i.e. the one from which the user arrived to this recipe, is indicated by a coloured background, helping the user to keep track of where he or she was going.

ondly, it allows a direct and effortless transition between the diplomatic and reading presentations of a recipe.



# Appendix E

## XSL Transformations

This section describes the XSLT scripts used to produce the different renditions of the edition, where to find them on the CD-ROM and gives some pointers for adapting them for more general use.

### E.1 Generating the diplomatic transcriptions

The diplomatic transcriptions that make up appendix B have been produced by applying a series of 23 eXtensible Stylesheet Language (XSL) Transformations, included in folder Appendix E - XSL Transformations/Diplomatic, to the base data files of the six manuscript versions. The conversion process was divided into five principal stages:

- 1) Extracting individual pages based on the milestone page markers
- 2) Enclosing text columns within elements based on the milestone column markers
- 3) Converting the pages from TEI XML to XHTML
- 4) Compiling textual notes into footnotes
- 5) Postprocessing for presentation

The first of these stages is performed by a single XSL Transformation, `1_Extracting_pages.xsl`, which was applied to the base data files of all six manuscript versions. This transformation processes each file and extracts all individual pages based on the milestone elements indicating page changes, ensuring that elements crossing page boundaries are properly divided and nested into a single `<div>` element representing the page. These individual pages are output into separate XML files, each containing the XML representation of the page and the full TEI header for the manuscript version. All of the remaining stages are applied to these individual page documents by chaining together the remaining 16 XSL Transformations.

The second stage applies only to pages which contain multiple columns and consists of the following transformations:

**2a\_Inserting\_column\_blocks.xsl** Replaces the milestone `<cb>` elements with enclosing `<div>` elements of `@type 'column'` and surrounds those column el-

ements with a `<div>` of `@type 'multicolumn'` by using straightforward string replacement instead of normal XSLT element insertion, with safeguards to deal with empty columns.

**2b\_Cleaning\_column\_blocks.xsl** Removes empty text-structural divisions that can be left at the beginning of a multicolumn page, moves the contents of columns spanning the entire width of the page outside of the multicolumn `<div>` and removes all column divisions spanning the entire page.<sup>1</sup>

**2c\_Fixing\_id\_numbers.xsl** Adds a subdivision identifier to all parts of elements that cross column boundaries and have thus been divided into several parts in order to avoid duplicate `@xml:id` values.

**2d\_Moving\_marginal\_notes\_outside\_columns.xsl** Moves displaced items that are located in the top or bottom margin of a multicolumn page to the beginning or end of the page outside the multicolumn block so that they can be correctly placed in the CSS representation using CSS.

After each page has been separated into its own file and columns have been marked with enclosing elements, each page is converted to XHTML by the main transformation, `3a_HTML_conversion.xsl`. This relatively complicated transformation consists of over 10,000 lines of code, most of which deals with the automatic generation of textual notes describing damaged, unclear or missing text and various types of emendation.<sup>2</sup> In addition to the straightforward replacement of the TEI elements with mostly `<span>` and `<div>` elements of various `@class` attribute values, this transformation also performs various adjustments and optimisations to facilitate the formatting of the pages using CSS. The main conversion script is followed by a small cleanup script (`3b_Removing_spaces_from_graphics.xsl`) that removes extra whitespace introduced by the conversion to the SVG structures used to replace the `<graphic>` elements representing various kinds of ornamental figures.

The fourth stage involves the combination and adjustment of the procedurally generated footnotes in order to minimize the number of footnote markers and avoid the duplication of footnotes with identical content. Many of the operations involved in this would be extremely complicated to implement in a single pass, which is why this process has been divided into sequential steps, performed by the following nine transformations:

**4a\_Joining\_footnoted\_spans.xsl** Joins together any consecutive `<span>` elements (and their enclosing footnote references) representing damaged or unclear passages that are contained within `<a>` elements referring to footnotes with identical content in order to eliminate unnecessary footnote pointers between two identically annotated segments.<sup>3</sup>

<sup>1</sup> Since columns are annotated using empty milestone elements, their ends cannot be explicitly indicated, and the end of a column needs to be inferred from the presence of a page break or a new column break—the beginning of a column spanning the full width of the page indicating the end of the multicolumn layout.

<sup>2</sup> Some of the complexity and length of the transformation is also explained by the fact that it was written in a relatively early stage of the editorial process with relatively little experience of XSLT, which means that it could most likely be rewritten in a more efficient and less verbose way.

<sup>3</sup> At this stage the footnote contents are still located within the `<a>` element at their point of reference, enclosed within a `<p>` element.

- 4b\_Joining\_contained\_spans.xsl** Similarly to the previous one, joins together `<unclear>` elements with identical attributes that were located within spans combined in the previous step.
- 4c\_Combining\_nested\_footnotes.xsl** Combines the contents of multiple nested footnotes that cover the same textual segment into a single footnote.
- 4d\_Adding\_unclear\_notes.xsl** Appends a note describing the extent of illegibility to damaged or deleted segments which have been rendered wholly or partially unclear by the damage or deletion.
- 4e\_Combining\_unclear\_notes.xsl** Incorporates the contents of notes describing individual unclear segments within damaged or deleted passages into the note describing the surrounding damage or deletion.
- 4f\_Numbering\_footnotes.xsl** Provides sequential numbers to the first instance of each unique footnote to be used as pointers in the text; notes whose content is identical to an earlier footnote will inherit the pointer from these in the next stage.
- 4g\_Consolidating\_footnotes.xsl** Moves all unique footnotes to a special section at the end of the file, adds pointers to the text and removes the content of the footnotes from the text.
- 4h\_Moving\_footnote\_spaces.xsl** Moving any spaces left at the end of `<span>` and `<a>` elements representing footnotes to the outside of these elements in order to bring the footnote pointers closer to the words or phrases they refer to.
- 4i\_Marking\_nested\_textnotes.xsl** Explicitly identifies all `<a>` elements representing links to footnotes that have another footnote link as their last child element and those footnote `<a>` elements which are the last child of another footnote element in order to allow the correct positioning of their pointers using CSS.

The final postprocessing stage involves various operations that have to do with the visual presentation of the manuscript pages in a browser. The transformations used for this stage are divided into three groups, the first four having to do with fitting the text onto the digital page, the fifth with the correct positioning of damage on displaced elements, the sixth with the rendering of different kinds of deletion and other markers using SVG, and the seventh with formatting the sentence structures of procedurally generated footnotes:

- 5a\_Marking\_long\_lines\_in\_columns.xsl** Marks those line break elements within multicolumn layouts, whose following content is longer than a specified threshold, representing the upper limit of the amount of text that can be expected to fit to the column.
- 5b\_Enclosing\_long\_lines\_in\_elements.xsl** Surrounds all elements that are contained on long lines within multicolumn layouts by an enclosing element so that they can be compressed using the *transform* property introduced by CSS3 to avoid visual overlap.
- 5c\_Consolidating\_long\_line\_elements.xsl** Combines consecutive elements marking compressed content into one element to avoid extra space between consecutive transformed elements.
- 5d\_Compressing\_displaced\_items.xsl** Handles cases where long lines in multicolumn layouts contain displaced items.

**5e\_Moving\_damage\_and\_notes\_into\_displaced\_items.xsl** Breaks up any damaged (and unclear) spans that contain displaced items—as well as their associated footnotes—at the boundaries of the displaced items and moves them to within the displaced elements to allow them to be rendered in the correct place.

**5f\_Rendering\_text\_decoration.xsl** Places various kinds of SVG figures representing different kinds of strikethrough, subpunction and other markers of deletion on top of text segments, scaling them dynamically based on the quantities and relative widths of the characters included in the segment, and adjusts various other things for a more accurate visual representation of the page elements.

**5g\_Formatting\_generated\_notes.xsl** Reformulates many of the footnotes that are the result of combining together several automatically generated notes and thus consist of several independent clauses into longer and more fluid sentences.

After having passed through this series of transformations, this XHTML file is styled using a CSS stylesheet specific to each manuscript version. These stylesheet files are generated by the transformation `Creating_diplomatic_CSS_for_print.xsl` which is applied to the appropriate base data file and extracts various metadata from the TEI header to scale the layout of the digital representation to approximate the original document and to define coloured representations for the different hands used in the manuscript. To produce the PDF representation included in appendix B from the HTML files, they were then batch processed by *wkpdf* and embedded into the appendix document using the *pdfpages* package for  $\text{\LaTeX}$ .

## E.2 Generating the parallel presentation

The parallel presentation included in appendix C was generated from the base data files and several of the annotation overlays by first applying the XSL Transformations `Parallel/Creating_parallel_reading_edition.xsl` and `Parallel/Rendering_explanatory_notes.xsl` to the annotation overlay file `PD_recipe_versions.xml` (see appendix A). The following files (included in appendix A) need to be present in the same folder with the transformation:

```
MSDouce55.xml
MSCosinViii11A.xml
MSHarley279.xml
MSAshmole1439.xml
MSAdditional5467.xml
MSHarley4016.xml
MSDouce55_expansion.xml
MSHarley4016_expansion.xml
MSHarley279_expansion.xml
MSCosinViii11A_expansion.xml
MSAshmole1439_expansion.xml
MSAdditional5467_expansion.xml
PD_recipe_sources.xml
```

PD\_recipe\_versions.xml

PD\_canonical\_titles.xml

This transformation builds an XML representation of all the unique *PD* recipes organised into the groups identified in chapter 13. The XSL Transformation `Parallel/Removing_leading_spaces.xsl` was then applied to the resulting file after whitespace normalization to clean away extraneous whitespace at the beginnings of headings and paragraphs. The resulting XML fragment was then inserted to the source file for appendix C (found on the CD-ROM). The tables of recipes for the six MS versions and the reading versions of the bills of fare included in the parallel reading edition were generated by applying the transformations `Creating_recipe_tables.xsl` and `Creating_parallel_menus.xsl`, respectively, to each of the base data files in turn and inserting the resulting XML fragments to the appendix source file. Finally, the XSL Transformation `latexPhD_appendix.xsl`<sup>4</sup> was applied to the source file to produce a LaTeX source file that was then typeset using Xe-LaTeX to produce the PDF output found in the /Appendix C - Parallel Reading Edition folder on the accompanying CD-ROM.

## E.3 Generating the HTML edition

Since the HTML edition combines the features of the diplomatic and parallel reading editions, it was produced in quite a similar way. The diplomatic representations included in the HTML were produced using exactly the same scripts described above for the printable diplomatic transcripts, the only difference being that a parameter named 'digital\_version' in the `3_HTML_conversion.xsl` transformation was set to the value 'true()', causing navigation hyperlinks to be added to each page.

Since the logical structure of the digital parallel reading edition is completely different from the printable one, being a network of interconnected files instead of a sequence of recipes, also the script used to produce the files making up the network—`HTML_parallel_edition.xsl`—is very different from one used for the static version. Because of the inclusion of the normalised form and word class in each word, using the stand-off annotation overlays involved constant file access operations, which are very costly in terms of processing time. In order to reduce the time required for the generation of the parallel reading files, an auxiliary pair of transformations—`Importing_expansions.xsl` and `Importing_normalisation+word-class.xsl`—were run on each base data file to import the expansions of abbreviated words and the normalised forms and word classes of all words into the base data file itself. These very simple scripts, which are included in the /Appendix E - XSLT Transformations/Overlays folder, also serve as a demonstration of the ease with which external metadata can be integrated to the base data file and exported again in a similar fashion.

With the linguistic metadata in the base data file, the transformation `HTML_parallel_edition.xsl` was applied to each base data file in turn, generating a separate table of contents file for each manuscript version, and a parallel presentation file

---

<sup>4</sup> This transformation is derived from the transformation used to convert this thesis itself into LaTeX for typesetting.

for every recipe in each six collections.<sup>5</sup> In addition to all of the source files, the use of this transformation also requires the annotation overlay files PD\_recipe\_versions.xml and PD\_canonical\_titles.xml to be in the same folder as the base data files.

---

<sup>5</sup> While the use of JavaScript or some other variety of dynamic scripting could have been used to significantly lower the number of files required, the completely static approach was chosen for the present purpose because of its robustness and ease of implementation, as no genuinely interactive interface was possible within the scope of this thesis.

## Appendix F

# Detailed contents of the *Potage Dyvers* MSS

This appendix contains detailed descriptions of the contents of all six *Potage Dyvers* manuscripts, expanding significantly on the outlines presented in chapter 9. The descriptions are based on both my personal examination of the original manuscripts and on earlier library catalogues and other sources, cited in the descriptions and listed in the main bibliography. The descriptions cover not only the recipe collection and any ancillary material (tables of contents and bills of fare), but in the case of the two miscellany manuscripts (MSS Ad and C), also all other works contained in the document.

Because of space considerations, the full descriptions have not been included in the printed version of this thesis, but are included as a separate appendix document on the enclosed CD-ROM (or .zip file in the case of the electronic edition). The descriptions are contained in a single PDF file located in the Appendix F - Manuscript Contents folder. The descriptions are based on the descriptions of the MS contents found in the <teiHeader> of each base data file in appendix A, but have been manually edited for more elegant print presentation.





## Appendix G

# LALME analysis data

In addition to the analysis results and data presented in part IV, the enclosed CD-ROM also includes full linguistic profiles for the six *PD* versions in PDF form as well as the fully quantified source files for these profiles in TEI XML form, along with some digital images of the distribution of individual *LALME* questionnaire items. This data is provided to allow the evaluation and verification of the results reported in part IV on the one hand, and to provide material for further analyses beyond the initial ones undertaken in this thesis. The different kinds of analysis data are contained in subdirectories of the */Appendix G - LALME Data* folder.

### G.1 Linguistic profiles

In order to facilitate the visual comparison of the linguistic profiles, the */Appendix G - LALME Data* folder contains a PDF document named *PD\_LALME\_Profiles.pdf*, which presents the linguistic profiles of the six *Potage Dyvers* manuscript versions—based on the questionnaire described in McIntosh, Samuels and Benskin (*LALME*: xviii-xxii)—side by side, listing all the questionnaire items which occur in at least one of the versions. The notation follows that used in *LALME* itself and described in McIntosh, Samuels and Benskin (*LALME*: vol.3: xiv). Forms whose frequency is at least two thirds of the frequency of the most frequent item are considered dominant and shown without parentheses, while forms whose frequency is between one and two thirds of the frequency of the most frequent item are shown in single parentheses, and minor forms whose frequency is less than one third of the frequency of the most frequent item are shown in double parentheses. Forms whose frequency is less than one per cent of the most frequent item are considered unlikely to be intentional and omitted from this table.

Although the specialised lexis and formulaic syntactic structure of the recipes mean that many of the items in the *LALME* questionnaire never occur in the recipes, it was nevertheless possible to collect reasonably detailed linguistic profiles of the texts. The category of features that suffered most noticeably from the syntactic peculiarities of the recipe text type is that of verb forms. The fact that the *LALME* questionnaire only covers indicative verb forms—although this is not very clearly documented—and does not separately register imperative or subjunctive forms—

constituting the majority of verb forms in recipe texts—means that the frequency of verb forms registered in the questionnaire is quite low. Since imperative and subjunctive forms differ significantly from the indicative forms, they are not included in the person-specific verb sub-forms,<sup>1</sup> although they *are* included in items which record all verb forms (label *vb*) or all present forms of the verb (label *pres*). While the former item is taken to encompass all forms of the verb, inflected or otherwise, the second is understood as representing all those present tense forms not falling under the person-specific sub-items; this has not been documented very clearly in the *LALME*, but the lists of forms for these items would seem to support this, as for example the second-person singular forms (or even their stems) do not seem to appear among the forms listed for *pres* items.

## G.2 Fully quantified linguistic profiles

The folder */Appendix G - LALME Data/Quantified LALME profiles* on the CD-ROM contains six TEI XML files, each containing an entry for each item of the *LALME* questionnaire for which forms were retrieved from the transcriptions of each of the six manuscript versions. Each entry takes the form of a `<div>` element formatted as follows:

---

**XML Example 19:** *An example of a quantified entry for a LALME questionnaire item.*

---

```
<div n="41">
  <head type="item">%WHILE</head>
  <list type="simple">
    <item><w>while</w><measure quantity="16" unit="count"></measure></item>
    <item><w>whill<ex>e</ex></w><measure quantity="5" unit="count"></measure></item>
    <item><w>whille</w><measure quantity="4" unit="count"></measure></item>
    <item><w>whyte</w><measure quantity="1" unit="count"></measure></item>
  </list>
</div>
```

---

These quantified profiles were produced by first creating ‘integrated’ versions of the base data files by importing the expansions of abbreviated words and the normalised forms and word classes of all words into them.<sup>2</sup> In order to extract the raw data for the profiles, the *Extracting\_LALME\_forms.xml* XSLT script—found in the same folder as the profiles themselves—was run on each of the ‘integrated’ base data files. This raw data—which contained not only the quantitative results for each item, but also KeyWord-In-Context (KWIC) concordances of all potential occurrences for those items whose forms could not be located with complete certainty—was then analysed and edited manually to determine the number of genuine instances of each form for each item, and to produce the quantified profiles included on the CD-ROM. While these quantified profiles are not compatible with the data presented in *LALME* itself, they have been included in the name of transparency.

<sup>1</sup> Neither are they treated as infinitives, although the imperative and subjunctive forms of most verbs are identical to it in present-day English.

<sup>2</sup> This was accomplished by running the *Importing\_expansions.xml* and *Importing\_normalisation+wordclass.xml* scripts (found in the */Appendix E - XSLT Transformations/Overlays* folder) on the original base data files.

## G.3 Exclusion templates for individual LALME forms

The folder /Appendix G - LALME Data/LALME item maps contains two digital image files for each individual dialectal form, based on data obtained using the *User-defined Maps* function of eLALME (Benskin, Laing et al. 2013) and used for the analyses presented in chapter 12. The first category of images, found in the Narrow subfolder, combines a *dot map* that visually represents all of the individual LALME LPs that exhibit the form in question with the more conservatively defined *exclusion template* that is used for the graduated exclusion templates (Figures 12.3, 12.5, 12.8, 12.10, 12.12 and 12.14). The second category of images, found in the Wide subfolder, consists of the more inclusive exclusion templates used for the absolute exclusion templates (Figures 12.2, 12.4, 12.7, 12.9, 12.11 and 12.13). The difference between the two forms of exclusion templates is that the latter one uses a larger radius for defining the area for which the form is considered to be typical and includes in it also ‘islands’ that do not exhibit the form but are completely surrounded by areas exhibiting it. The image files are named using the formula [MS\_ID]\_[LALME\_item\_no]\_[form(s)].png.<sup>3</sup>

---

<sup>3</sup> The fact that the same forms occur in multiple manuscripts means that there are multiple copies of some images, but this redundancy has been preserved in the name of consistency and ease of finding them.