Mémoires de la Société Néophilologique de Helsinki

édités par Juhani Härmä, Jarmo Korhonen et Terttu Nevalainen

Tome XCIV

ISBN 978-951-9040-51-6 (PDF)

Mémoires de la Société Néophilologique de Helsinki
Tome XCIV

Sociolinguistic variation in English derivational productivity

Studies and methods in diachronic corpus linguistics

by
Tanja Säily

Academic dissertation to be publicly discussed, by due permission of the Faculty of Arts at the University of Helsinki, in the Small Hall on the 31st of October, 2014, at 12 o'clock.

Helsinki 2014
Société Néophilologique

ISBN 978-951-9040-50-9
ISSN 0355-0192
Hansaprint
Vantaa 2014

Preface

In the spirit of the digital humanities, this work is a product of multidisciplinary collaboration. As a computer geek and an ex (or eternal) engineering student, I gained an interest in corpus linguistics soon after beginning my studies of English Philology at the University of Helsinki. I got the idea of studying sociolinguistic variation in morphological productivity from Professor Terttu Nevalainen, who taught the advanced studies seminar on sociolinguistics and who was to become the supervisor of both my MA and my PhD theses. I got the initial material – and the training to use it – from her *Corpus of Early English Correspondence* team. I discussed the problem of comparing type counts with my significant other, computer scientist Jukka Suomela, who promptly came up with a version of the method I have used and developed throughout this work.

Terttu and Jukka (the latter now a professor at Aalto University) are the two people to whom I owe the greatest debt of gratitude. Thank you for your hard work in mentoring, assisting and co-authoring papers with me, as well as for your infinite patience, contagious enthusiasm, great ideas, and freely shared and apparently fathomless knowledge of your respective fields. And, you know, thanks for sharing your life with me, Jukka.

But wait! There is more. The beginning of my PhD studies in 2009 was also the beginning of our Academy of Finland funded project, *Data mining tools for changing modalities of communication* (with the rather Shakespearean-insult-like acronym DAMMOC), with members from the University of Helsinki, Aalto University and the University of Tampere. My other two co-authors came from this project: the brilliant Dr. Jefrey Lijffijt from Aalto, who earned his doctorate in our project, and Dr. Harri Siirtola from Tampere, whose mad skillz in information visualisation continue to amaze me.

The PI of the project, Professor Heikki Mannila from Aalto (now President of the Academy of Finland), always had excellent feedback and questions. I would also like to thank Professor Panagiotis Papapetrou and Dr. Kai Puolamäki for comments, discussions and all around good company. Many thanks to Professor Kari-Jouko Räihä from Tampere for his contributions to our visualisation papers. The Helsinki members of the project were Professor Nevalainen and myself as well as Mr. Turo Vartiainen, our hard-core linguist, a fount of knowledge and spreader of good cheer. I am delighted to continue my collaboration with Terttu, Turo, Jefrey and Jukka in a new project entitled *Reassessing language change: the challenge of real time* (acronym to be announced).

I have been fortunate to be able to conduct my research under the auspices of the Research Unit for Variation, Contacts and Change in English (VARIENG). My first contact with the unit was as a research assistant to Dr. Roderick McConchie, who was working on word-formation and the proceedings of the HELLEX symposium, which resulted in my first academic publication as a co-editor. Then I began my training in the CEEC team, ably headed by Professor Nevalainen and Dr. Helena Raumolin-Brunberg, receiving an especially great deal of feedback and assistance from Dr. Arja Nurmi and Mr. Samuli Kaislaniemi, with both of whom I have had the pleasure to share an office. My successor as research assistant, Mr. Mikko Hakala, helped us with the noun–pronoun paper and quickly became an invaluable asset, friend and colleague. Heartfelt thanks also to other CEECers past and present: Professor Minna Palander-Collin, Professor Mikko Laitinen, Dr. Minna Nevala, Dr. Anni "Blades of Glory" Sairio, Mr. Teo Juvonen and Ms. Tuuli Tahko.

Other past and present Variengers, too, are to be thanked for creating a genial and intellectually stimulating working environment: Dr. Alexandra Fodor, Dr. Alaric Hall, Dr. Turo Hiltunen, Dr. Marianna Hintikka, Dr. Alpo Honkapohja (our supplier of Swiss chocolate), Dr. Leena Kahlas-Tarkka, Dr. Matti Kilpiö, Professor Martti Mäkinen, Dr. Dr. Ville Marttila, Dr. Anneli Meurman-Solin, Professor Päivi Pahta, Dr. Tiina Räisänen, Dr. Maura Ratia, Professor Emeritus Matti Rissanen, Dr. Maija Stenvall, Dr. Carla Suhr (#crossbow), Professor Emerita Irma Taavitsainen, Professor Olga Timofeeva, Dr. Heli Tissari, Professor Jukka Tyrkkö (aka Salient Jukka), Dr. Irmeli Valtonen, Dr. Anna-Liisa Vasko, Mr. Simo Ahava, Mr. Henri Kauhanen, Ms. Minna Korhonen (also a pleasant office mate), Ms. Anu Lehto, Mr. Joe McVeigh (my eSeries colleague), Ms. Sara Norja, Ms. Raisa Oinonen, Ms. Ulla Paatola, Mr. Mikko Alapuro (who kept the CEECer software going), Ms. Emanuela Costea, Ms. Anne Kingma (who found a missing reference) and Ms. Tuula Nikander. Many a good time has been had by the varijrs over the years.

I am grateful to the Jyväskylä contingent of VARIENG, led by Professor Sirpa Leppänen, for refreshingly different views on research into the English language. Special thanks go to Dr. Samu Kytölä and to fellow blog committeers Dr. Alicia Jinkerson, Ms. Saija Peuronen and Dr. Elina Westinen.

I wish to thank Langnet, the Finnish doctoral programme in language studies, for four years of funding as well as for the exciting seminars of the Language Variation and Change subprogramme, which was first led by Professor Taavitsainen and Professor Ulla-Maija Forsberg, then by Professor Riho Grünthal and Professor Juhani Klemola, all of whom I thank for their support and interest.

During my Langnet term, I broadened my horizons thanks to Ms. Riikka Ala-Risku, Ms. Sonja Dahlgren, Mr. Ludvig Forsman, Ms. Lotta Jalava, Mr. Santeri Junttila, Ms. Hanna Lantto, Ms. Heini Lehtonen, Dr. Kaarina Mononen, Dr. Paula Rautionaho, Ms. Zsuzsanna Renkó-Michelsén, Ms. Iiris Rennicke, Ms. Mari-Liisa Varila and Mr. Max Wahlström. I visited the Hermann Paul School of Linguistics in Freiburg in 2010, discussing research and other things with many amiable people in the Frequency Effects in Language group, including Dr. Florent Perek, Dr. Ulrike Schneider, Dr. David Lorenz and Dr. Malte Rosemeyer.

I have also benefited from attending conferences, where I have been able to discuss my work with scholars like Professor Harald Baayen, whom I thank for his helpful comments. I got the idea of using interactive SVG images embedded on web pages from Dr. Jack Grieve at AVML 2012. I am grateful to Dr. Claire Cowie, Dr. Anne Gardner, Mr. Mark Lindsay, Dr. Cynthia Lloyd, Ms. Marta Lupica Spagnolo, Professor Anne Schröder and Professor Pavol Štekauer for sharing their work on word-formation with me. Special thanks to Dr. Gardner for using our method and for many discussions at various occasions, most recently over a memorable dinner at ISLE 3.

I would like to thank the pre-examiners of my dissertation, Professor Martin Hilpert and Professor Antoinette Renouf, for helpful and encouraging feedback. Professor Hilpert has also kindly agreed to act as my opponent, for which I am grateful in advance. I thank the Modern Language Society for publishing this dissertation in their *Mémoires* series, and Dr. Hintikka, Ms. Marja Ursin and Ms. Tiina Kaarela for their help during the final stages.

Finally, I would like to thank my (non-work-related) friends and family for their support. My crafty friends, chief among them Raisa Asikainen, have been instrumental in much needed getaways devoted to artistic pursuits. My family, as well as Jukka's, have been nothing but supportive during my extended adventures in academia. I have partly inherited my research interests from my mother, Riitta, who is a retired teacher of English and Swedish, and my father, Markku, a retired science teacher who wrote some of the first Finnish textbooks on automatic data processing and information technology. My brother, Tatu, and I were computer geeks from the first Compis system we had at home. Thank you all for letting me do my thing. Now the thing is a book, and the adventure continues.

Helsinki, October 2014 *Tanja Säily*

Contents

List of tables

List of figures

PART I: BACKGROUND

1. Introduction

1.1. Motivation

The study of derivational productivity has for a long time concentrated on structural constraints and, since the 1990s, variation across genres. Even though sociolinguistic variation was hypothesised to underlie changes in productivity decades ago (Romaine 1985), it is only in the past few years that scholars have begun to conduct corpus-linguistic research into sociolinguistic variation and change in productivity. While part of the reason for this has been the lack of suitable and easily available corpora designed for sociolinguistic research (cf. Kendall 2011), another serious issue has been the lack of suitable and easily available methods. This dissertation develops such a method and uses it to study sociolinguistic variation in the productivity of the nominal suffixes *-ness* and *-ity* in Early Modern, Late Modern and Present-day English material.

Until recently, diachronic corpus linguistics in general has suffered from both scarcity of data and relatively unsophisticated methods in terms of statistical analysis and visualisation. With the growing availability of large historical corpora and tools for spelling standardisation, the methods used are becoming more advanced. However, with increasing sophistication comes increasing danger of a black-box analysis entailing background assumptions that may not be clear to the researcher and that may affect the reliability of the results. Therefore, the present work advocates robust, data-driven methods which make no simplifying assumptions about the data. Furthermore, this work employs state-of-the-art techniques from information visualisation that facilitate exploration and the discovery of outliers. The methods benefit users of large and unstructured corpora as well as those of smaller and more carefully compiled corpora.

1.2. Setting the scene for *-ness* and *-ity*

The two suffixes studied in this work are the roughly synonymous *-ness* and *-ity*, which are typically used to form abstract nouns from adjectives, with the approximate meaning 'the state or quality of being ADJ'. Thus, given an adjective like *productive*, we can use either of the suffixes to form an abstract noun, i.e., *productiveness* or *productivity*. Nevertheless, the suffixes differ in the kinds of bases and genres they prefer, the functions in which they are used and, arguably, in their semantics (see Section 2.4 below). This is due to their history: while *-ness* is a native suffix, *-ity* came into Middle English from French borrowings and was later reinforced through calques on Latin. It is thus to be expected that the use of

the suffixes also varies sociolinguistically, especially in the case of the more learned and prestigious *-ity*. However, very little research has been conducted on sociolinguistic variation in their productivity, even though the pair has served as an example of competing affixes in multiple studies. Productivity can be defined as "the statistically determinable readiness with which an element enters into new combinations" (Bolinger 1948: 18; see further 2.3 below).

To set the scene for the suffixes, let us consider sociocultural and stylistic trends that influenced the use of the English language in England during the periods studied here, i.e., the 17th, 18th and late 20th centuries (cf. Culpeper and Nevala 2012). While many of the trends cut across several periods, they are here considered under the period in which they could be said to have culminated.

The Renaissance, which in the English context extended into the **17th century**, saw the rise of English as a national language modelled on the Latin which it displaced (Adamson 1999: 541ff.). The goal of grammar school was to make the pupils (who were boys from the higher social ranks) bilingual in English and Latin, and good style was defined as a learned, Latinate style. There was wisdom and power in eloquence, especially in terms of *copia*, or the abundance of words, which included the copious use of synonyms and morphologically related forms. This seems to have led to a peak in ephemeral borrowings and native formations in the 16th century (Nevalainen 1999a: 349). By the mid-17th century, however, copious borrowing from Latin was no longer in vogue, and authors began to speak in favour of native means of achieving a classical style. Nevertheless, Adamson (1989: 214) argues that the native and Romance/Latinate resources have remained in English as separate stylistic strata with connotations of "physical reality and subjective response" and "conceptual clarity and emotional neutrality", respectively.

Rather than *copia*, the ideal of the **18th century** was one of perspicuity (Adamson 1999: 599ff.). English style was to be "familiar, but not coarse, and elegant, but not ostentatious" (Adamson 1999: 615, citing Samuel Johnson), although poetry was allowed a grander vocabulary. The popular genres of the 18th century temporarily reversed the trend of increasingly situation-dependent reference between 1650 and 1990, preferring more elaborated reference (Biber and Finegan 1997). Furthermore, the standardisation process of the English language reached the stage of prescription during the course of the 18th century (e.g., Nevalainen and Tieken-Boon van Ostade 2006). All aspects of language were codified in grammars, which began to formulate a norm of correct English. The belief in a single correct standard is a persistent one even in today's society.

The **late 20th century** was a time of continued scientific and technological innovation (Romaine 1998). New scientific and technical terms were predominantly formed using Greek and Latinate elements, which were highly productive in this specialised field. Other ongoing trends included urbanisation and democratisation, the latter of which has been seen as the cause of the colloquialisation of several genres of written English in the 20th century (Hundt and Mair 1999). However, many specialist genres have been diverging in the opposite direction since the 18th century (Biber and Finegan 1997).

1.3. Research questions

The major issues addressed in this dissertation can be summarised in the following research questions.

1. Is there sociolinguistic variation and change in the productivity of *-ness* and *-ity* in the history of English?

Question 1 is the starting point of my work. There is very little previous research into word-formation from the point of view of sociolinguistics, let alone historical sociolinguistics (see Section 2.3.4.3 below). This pair of suffixes is well-studied in terms of purely linguistic factors, but despite the fact that they are known to belong to different stylistic strata, the sociolinguistic factors affecting their use have thus far been mostly ignored, with the exception of a few qualitative remarks (see 2.4 below). A thorough corpus-linguistic investigation is therefore in order. The question is probed using corpora representing various genres and time periods. Chapter 6 focuses on 17th-century letters, Chapter 9 analyses late 20th-century data, Chapter 10 compares 18th-century letters and courtroom discourse, and finally, Chapter 11 takes a closer look at 18th-century letters, trying to determine whether the change in the productivity of *-ity* observed in the 17th and 18th centuries is linguistic or stylistic.

While the studies are exploratory in nature, two research hypotheses are formed in advance. Firstly, the productivity of *-ity* is assumed to be significantly low in letters written by women and the lower social ranks in the 17th century, because *-ity* is an etymologically foreign suffix that may initially only have been available to people who had a classical education, who were most often high-ranking men. This hypothesis is later modified based on the results of the first two studies, but *-ity* is still considered to be susceptible to variation. Secondly, the productivity of *-ness* is assumed to be basically invariant, as previous research has seen it as the default suffix for forming abstract nouns from adjectives. In

addition to gender and social rank, the social categories used in the exploratory analysis include register in the sense of participant relations; all of these are discussed further in Chapter 3 below. Another factor explored is change over time.

2. How can we study productivity in small corpora which contain a great deal of spelling variation?

Question 2 is a methodological issue that arises immediately at the outset of the first study. There are no large historical corpora compiled according to sociolinguistic principles at the moment; a similar problem is faced by, e.g., studies of Old English, where the largest corpus contains all there is left of the language, which is around three million running words (see CoRD). The small amount of data limits the kinds of analysis we can conduct as well as the reliability of the results.

In materials produced before the advent of standard language, there is also a great deal of spelling variation. While this issue can be alleviated somewhat by preprocessing the data using a program like VARD (Baron 2011; Baron and Rayson 2009), the spelling of rare words will still not be standardised automatically. This is problematic to studies of productivity, which are often not only concerned with rare formations but also with their bases, so the analyst should track the spelling variants of each of them. Chapter 6 presents a method for measuring morphological productivity in response to question 2.

3. How can we study variation and change in corpora which may not be completely comparable over time and across genres?

Both sociolinguists and historical linguists generally accept the uniformitarian principle, which states that people and their linguistic practices today are comparable to those in the past (Nevalainen and Raumolin-Brunberg 2012: 24–25). That is, the same mechanisms are seen to operate in both the present and the past (Labov 1978 [1972]: 161), which has enabled the application of sociolinguistics to historical material. As noted by Bergs (2012), however, this does not imply that social categories are invariant. Genres, too, change over time, and a key issue is the kinds of data available for different periods and genres (Cantos 2012). In order to make "the best use of bad data" (Labov 1994: 11), we need to be aware of how well the materials we are using match each other, both within and across corpora.

To this end, the present work includes two papers that present methods for comparing corpora and studying genre continuity. Chapter 7 does this by comparing word frequencies using a novel method for establishing statistical significance, while Chapter 8 focuses on comparing part-of-speech frequencies using innovative visualisations. The information provided by these studies is utilised in the chapters analysing variation and change in morphological productivity.

4. Are the productivity measures proposed in previous research valid in and applicable to sociolinguistic data of this kind?

Most of the previous research on productivity has concentrated on comparing the productivity of different affixes within the same corpus. To study variation and change in productivity, we need to compare the productivity of a single affix across subcorpora, which will usually vary in size. As discussed in Section 5.1.2.1 below, this is problematic because common measures of productivity depend on the size of the corpus in a non-linear manner, which means that normalising the frequencies is not an option. The new method proposed in Chapter 6 is assumption-free, highly visual and provides a built-in measure of statistical significance.

An important component of Baayen's measures of productivity (see 2.3.2 below) are hapax legomena, or words occurring only once in the corpus. Rather worryingly, Chapter 6 finds that in the 17th-century section of the *Corpus of Early English Correspondence*, the frequency of hapax legomena is not a suitable measure for the productivity of *-ness* and *-ity*, as the frequency in a given subcorpus seems to be largely a matter of chance. Therefore, Chapter 9 studies the theoretical and empirical validity of hapax-based measures of productivity in sociolinguistic research.

5. What are the requirements for a usable tool for studying variation in productivity in data of this kind?

Anthony (2013) argues that as corpus linguistics matures, there is a growing need for more sophisticated tools, which can in practice only be developed by collaborating with members of the science and engineering community. This is exactly what I have been doing for the past few years. In the course of my work on productivity, I have encountered a number of issues with the way that the method presented in Chapter 6 was originally implemented (Suomela 2007). Chapter 10 discusses these issues and presents a new version of the implementation (Suomela 2014), which provides several improvements.

To clarify the relationship between the research questions and the studies, Table 1.1 provides an overview of which questions are answered in which chapters.

Table 1.1. The relationship between the chapters and the research questions

	Chapter					
	6	**7**	**8**	**9**	**10**	**11**
RQ 1	•			•	•	•
RQ 2	•					
RQ 3		•	•			
RQ 4	•			•		
RQ 5					•	

To conclude this section, I wish to point out that most of my research questions have in fact been formulated and answered in collaboration with others. I have been fortunate to work with a team of linguists and computer scientists in the DAMMOC project, funded by the Academy of Finland Motive programme in 2009–2011. My work on productivity has been done in a long-term collaboration with Jukka Suomela, another computer scientist. Nevertheless, I can confidently say that without my contribution, the studies presented in this dissertation would not exist, and the methods herein would not have been introduced into diachronic corpus linguistics. The next two sections describe the structure of the dissertation and the provenance of the studies, after which Section 1.6 lists my contributions to the studies.

1.4. Structure of the dissertation

The contents of this work are organised as follows. Comprising Chapters 1–5, **Part I** presents the background to the research, beginning with the introductory remarks in this chapter. Chapter 2 considers theoretical and methodological issues in word-formation and productivity, finishing with previous research on *-ness* and *-ity*. Chapter 3 shifts the focus to historical sociolinguistics by discussing theoretical and methodological issues to do with the linguistic variable as well as the social categories studied in this dissertation, embedded in the sociohistorical contexts of the material used. The material receives its own introduction in Chapter 4, with especial attention paid to the amount of data available in the different (sub)corpora. Chapter 5 presents the most important and innovative

methods used in the studies, relating them to state-of-the-art methodology in diachronic corpus linguistics.

Part II consists of the six studies that together form the chief contribution of this dissertation, designed to answer the research questions posed in Section 1.3 above. Chapter 6 explores sociolinguistic variation and change in the productivity of *-ness* and *-ity* in 17th-century correspondence, introducing a method for analysing productivity variation across subcorpora. Chapter 7 investigates genre continuity and lexical change in 17th-century correspondence and presents two methods for comparing corpora. Chapter 8 studies genre continuity and sociolinguistic variation in noun and pronoun frequencies in early correspondence, using beanplots to visualise the differences. Chapter 9 focuses on sociolinguistic variation in productivity in Present-day English and analyses the validity of hapax-based measures of productivity. Chapter 10 examines sociolinguistic variation and change in productivity in 18th-century English, proposing improvements on the method introduced in Chapter 6. Using the improved method, Chapter 11 zooms in on sociolinguistic variation and change in productivity in 18th-century correspondence, considering factors such as register, semantics and the influence of normative grammar.

Part III brings the dissertation to a close by discussing the studies presented in Part II. Chapter 12 provides a summary of the answers to the research questions asked in Section 1.3. Chapter 13 critically evaluates the main methods used in the studies, beginning with morphological productivity and proceeding to statistical significance and visualisation. Chapter 14 evaluates and explains the key findings on *-ity* and *-ness*, linking them to previous and future research. Finally, Chapter 15 concludes the dissertation by considering its implications for the future in terms of *-ness* and *-ity*, variation in morphological productivity and corpus-linguistic methodology. To assist the reader, two appendices are provided: (I) a glossary of statistical terms and (II) a list of the chief sociolinguistic parameters of the *Corpora of Early English Correspondence*.

1.5. Provenance of the studies

Chapters 6–11 in Part II have been published or accepted for publication in the following peer-reviewed sources and are printed in this dissertation with the permission of the publishers.

Chapter 6:
Säily, Tanja and Jukka Suomela 2009. Comparing type counts: The case of women, men and *-ity* in early English letters. Antoinette Renouf and Andrew

Kehoe (eds.), *Corpus Linguistics: Refinements and Reassessments*, 87–109. Language and Computers: Studies in Practical Linguistics 69. Amsterdam: Rodopi.

Chapter 7:
Lijffijt, Jefrey, Tanja Säily and Terttu Nevalainen 2012. CEECing the baseline: Lexical stability and significant change in a historical corpus. Jukka Tyrkkö, Matti Kilpiö, Terttu Nevalainen and Matti Rissanen (eds.), *Outposts of Historical Corpus Linguistics: From the Helsinki Corpus to a Proliferation of Resources*. Studies in Variation, Contacts and Change in English 10. Helsinki: VARIENG. http://www.helsinki.fi/varieng/series/volumes/10/lijffijt_saily_nevalainen/

Chapter 8:
Säily, Tanja, Terttu Nevalainen and Harri Siirtola 2011. Variation in noun and pronoun frequencies in a sociohistorical corpus of English. *Literary and Linguistic Computing* 26(2): 167–188. Reproduced by permission of the European Association for Digital Humanities and the Alliance of Digital Humanities Organizations. doi:10.1093/llc/fqr004

Chapter 9:
Säily, Tanja 2011. Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory* 7(1): 119–141. Reproduced by permission of De Gruyter Mouton. doi:10.1515/cllt.2011.006

Chapter 10:
Säily, Tanja in press. Sociolinguistic variation in morphological productivity in eighteenth-century English. To appear in *Corpus Linguistics and Linguistic Theory*. (Special issue, ed. by Martin Hilpert and Hubert Cuyckens.) Reproduced by permission of De Gruyter Mouton.

Chapter 11:
Säily, Tanja forthcoming. Change or variation? Productivity of the suffixes *-ness* and *-ity*. To appear in Terttu Nevalainen, Minna Palander-Collin and Tanja Säily (eds.), *Change in 18th-Century English: New Approaches to Historical Sociolinguistics*.

1.6. Author's contributions

The first three studies in Chapters 6–8 are co-authored: I am the main author of Chapters 6 and 8, while Jefrey Lijffijt and I contributed equally to Chapter 7, with Terttu Nevalainen as the third author. Chapters 9–11 were authored by me alone.

In Chapter 6, I came up with the linguistic questions and hypotheses, retrieved the instances of *-ness* and *-ity* from the corpus and lemmatised them. I discussed the problem of comparing type and hapax frequencies with Jukka Suomela, who devised the method, programmed the software to realise it and created the figures, after which I analysed and interpreted the results. While we worked on the paper together, I wrote most of the *Introduction*, *Background and related work*, *Material* and *Results and conclusions*, whereas Suomela wrote most of the *Methods* and *Implementation* sections.

In Chapter 7, I developed the method of tracing changes in cultural vocabulary using the *Historical Thesaurus* and carried out the analysis of war-related words. Jefrey Lijffijt was responsible for the basic method of comparing word frequencies with the bootstrap test as well as implementing and applying it, while Terttu Nevalainen analysed linguistic changes. The experiments using the bootstrap test were co-designed by all three authors. I wrote the sections on *Approaching the Civil War effect in 17th-century England* (with the aggregated Google Books material provided by Lijffijt), *Data* and *The Civil War effect in vocabulary*, and we all co-wrote the *Introduction*, *Discussion* and *Conclusion*. The rest of the sections were written by my co-authors.

In Chapter 8, I came up with the linguistic questions and hypotheses together with Terttu Nevalainen. I analysed the tagging of nouns and pronouns in the corpus and designed the rules for reannotating the corpus with feedback from Nevalainen. The visualisations and statistical comparisons were carried out by Harri Siirtola, while I provided the sociolinguistic interpretation of the results. I wrote most of the text in the sections on *Issues with historical data*, *Material*, *Analysis of the reliability of POS tagging in the PCEEC*, *Discussion* and *Conclusion* as well as the appendices, and the *Analysis of shifts in tag ratios* was co-written by us all. The remaining sections were chiefly written by Nevalainen.

While I am the sole author of Chapters 9–11, I continued to develop and use the method for comparing type frequencies in collaboration with Jukka Suomela. He programmed the new version of the software described in Chapter 10 and utilised in Chapter 11. The linguistic side of the chapters is my own work, although I naturally received feedback on it from other scholars at various stages of the research.

2. Word-formation and productivity

To contextualise the studies of derivational productivity in Part II, this chapter introduces basic morphological concepts (Section 2.1), discusses the place of word-formation in theories of grammar (Section 2.2), examines the concept of morphological productivity from multiple viewpoints (Section 2.3) and surveys previous research on *-ness* and *-ity* (Section 2.4).

2.1. Basic concepts

Before embarking on a study of suffixation, some basic concepts need to be introduced and defined. The concept of a **word** I shall take *a priori*, but I shall split from it some more specific concepts. A **lexeme** comprises all the possible shapes that a word can have, such as *shoot*, *shoots*, *shooting* and *shot* for the verbal lexeme *shoot*; the individual shapes are called **word-forms** (Bauer 1983: 11). A **lemma** is the word-form conventionally used to represent a lexeme, e.g., in a standard dictionary (Bauer 1983: 12, who calls it a citation form). Especially in older material, there may be variation in the spelling of the word-forms; these variants, which are what we actually see in the text, can be called **orthographic forms**.

According to Bauer (1983: 13), **morphology** is the study of the internal structure of word-forms. As noted by Plag (2003: 10), a **complex word** like *unfaithfulness* (my example) can be broken down into its smallest meaningful units, **morphemes**: *un-*, *faith*, *-ful* and *-ness*. Plag (2003: 10) classifies morphemes into two kinds: **free morphemes** such as *faith* that can occur by themselves, and **bound morphemes** such as *un-*, *-ful* and *-ness* that can only occur with other morphemes. A free morpheme occurring by itself is called a **simplex** (Bauer 1983: 30) or **monomorphemic word** (Plag 2003: 25).

According to Plag (2003: 10–11), the central meaningful element of a word can be called the root, base or stem. Bound morphemes that attach to the central element are called **affixes**; these can be divided into **prefixes** (such as *un-*), which occur before the central element, **suffixes** (such as *-ness*), which occur after it, and **infixes** (such as *-bloody-* in *abso-bloody-lutely*), which occur inside it. Plag (2003: 10–11) explains the different terms for the central element as follows. The **root** consists of a single morpheme that can be either free like *faith* or bound such as the Latinate *simul-* (as in *simulant*, *simulate*, *simulation*). The **base** is a wider concept: it is used for any central element, whether an indivisible root or a complex word, to which an affix can be added. The **stem** has various meanings in

the literature, the most common of which is 'the base of an inflection'; following Plag's (2003: 11) lead, I shall avoid using this ambiguous term.

Plag (2003: 20–21) sees the morpheme as a linguistic sign that has two sides: form and meaning. For example, the morpheme *un-* consists of the form, or **morph**, [ʌn] and the meaning 'not'. The form of a morpheme can vary; these variants are called **allomorphs** (Plag 2003: 27–28). For instance, the form of the base *eccentric* [ɛk'sɛntrɪk] changes when the suffix *-ity* is attached to it: [ɛksɛn'trɪs]+[ɪtɪ] (Romaine 1985: 451). Plag (2003: 21) says that when two morphemes are combined, the meaning of the resulting complex word is often **compositional** and hence transparent – e.g., *un-* 'not' + *happy* 'happy' = *unhappy* 'not happy'. He notes (2003: 22), however, that this is not always the case – for example, *late* 'after the due time' + *-ly* 'in an X manner' = *lately* 'recently', not 'in a late manner' (see the discussion on lexicalisation below).

The **morphological process** of adding an affix to a base is called affixation (more specifically, prefixation, suffixation or infixation); this can be either inflectional or derivational. **Inflectional affixation** is used to create the different word-forms of a lexeme (Bauer 1983: 29); it encodes grammatical categories such as plural, person, tense or case (Plag 2003: 14). **Derivational affixation** is used to create new lexemes (Bauer 1983: 29), and it is a subtype of this process that I am concerned with here: creating new words by using *-ness* and *-ity* suffixation.

How do speakers form new words? According to Plag, word-formation is not an arbitrary process but seems to be **rule-governed**: for example, most adjectives can take the suffix *-ness*, and the resulting noun will regularly have the meaning 'the property of being X', where X denotes the meaning of the base (2006: 537). Or, given the words *unhappy*, *unkind*, *unfaithful*, *untrue*, *uncommon* and *analysable*, a speaker can easily decipher the meaning of *unanalysable*, even if she has not encountered that word before (Plag 2003: 30). There must be some kind of system in speakers' minds that makes this possible; according to Plag (2003: 37–38), some say it is the general mechanism of **analogy** that is at work, while others claim that when there are multiple instances of the same pattern, there must be a rule by which they are formed.

A typical word-formation rule might look like the one presented in (2.1), adapted from Plag (2003: 35).

(2.1) Word-formation rule *un-*$_1$

phonology:	/ʌn/-X
base:	X = adjective
semantics:	'not X'

constraints:	– derivatives with simplex bases must be interpretable as contraries – further restrictions on possible base words …

Analogy, on the other hand, is simply "a proportional relation between words", as exemplified in (2.2) below (Plag 2003: 37). In the first example, the relationship between items a and b is the same as the relationship between items c and d. Item d has been formed from c on the pattern of a : b. Concrete examples are provided in ii–iv.

(2.2) i. a : b :: c : d
ii. eye : eyewitness :: ear : earwitness
iii. ham : hamburger :: cheese : cheeseburger
iv. sea : sea-sick :: air : air-sick

In Plag's opinion (2003: 38), the advantage to a rule-based approach is that it explains the existence of systematic structural constraints on morphological processes as well as why some processes are more frequently utilised than others: the constraints are explicitly listed in the rule, and processes that are never or seldom used just do not have a rule associated with them (Bauer 2001: 77). However, as both Plag (2003: 38) and Bauer (2001: 96) admit, analogy is certainly employed to some extent; furthermore, I do not think that rules as clear as the one in (2.1) really exist in speakers' heads – the reality must be much fuzzier than that, with analogy playing a large part and interacting with other factors such as speakers' knowledge about how other speakers use the forms in question. The fuzziness hypothesis is supported by the considerable number of exceptions to the strict rules proposed by linguists (cf. Bauer 1983: 293–294).

In addition to fuzzy word-formation rules, speakers must have some words stored in their minds to which the rules can be applied. This storage space is called the **mental lexicon** (Plag 2003: 4). Words listed in the mental lexicons of speakers are called **existing words**, while words that are not listed there but could be formed by a rule are called **potential words** (Plag 2003: 46–47). Existing words can develop idiosyncratic meanings or pronunciations by a process known as **lexicalisation** (Bauer 2001: 44–45); the above-mentioned *lately* 'recently' is a case in point. Another good example is the word *business* ['bɪznɪs] 'the production of goods and services for profit', which has diverged in both form and meaning from the original ['bɪzɪnɪs] 'the state or property of being busy'.

The development of new word-formation rules can be seen as **grammaticalisation**, defined by Brinton and Traugott (2005: 99) as "the change whereby in certain linguistic contexts speakers use parts of a construction with a grammatical function. Over time the resulting grammatical item may become more grammatical by acquiring more grammatical functions and expanding its host-classes." The function could be, e.g., changing grammatical category (cf. 2.3.4.2 below). On the other hand, many rules include a semantic component, they have other functions beyond grammatical ones, and their use is restricted; Brinton and Traugott (2005: 91–92) regard the development of such a rule as lexicalisation.

Armed with these concepts, we may now dig deeper into the nature of word-formation and productivity.

2.2. Between lexis and morphology

As noted by Brinton and Traugott (2005: 91), "[a] major problem for theories of grammar has been where to locate word formation". Does it belong to morphology or to the lexicon? Brinton and Traugott (2005: 91, 96) themselves see productive word-formation as operating outside the lexicon; in their model, a productively formed word only enters the lexicon if it later undergoes the above-mentioned process of lexicalisation. To begin exploring this question further, let us consult three contemporary reference grammars of English.

A Comprehensive Grammar of the English Language (CGEL; Quirk et al. 1989 [1985]), described by Standop as "strukturalistisch–eklektisch" (2000: 248, as cited in Mukherjee 2006: 340), relegates word-formation to an appendix, albeit an extensive one. For Quirk et al. (1989 [1985]: 12), grammar consists of syntax and inflections, whereas word-formation inhabits the common ground between grammar and lexicology, grammar providing the rules and lexicology the idiosyncrasies (id.: 1517). While heavily influenced by the CGEL (see Mukherjee 2006), the *Longman Grammar of Spoken and Written English* treats both derivation and inflection under morphology and the general heading of "Word and phrase grammar" (Biber et al. 1999: Chapter 2). *The Cambridge Grammar of the English Language* goes so far as to dedicate a full chapter to "Lexical word-formation" (Huddleston and Pullum 2002: Chapter 19), which is seen as part of morphology and grammar, but "related to the dictionary" (id.: 28).

The most common answer, then, seems to be that word-formation belongs to both morphology and the lexicon in some way. The specifics of this are theory-dependent (Brinton and Traugott 2005: 91). Some theories (item-and-arrangement) take the morpheme as the basic unit of analysis, while others (item-and-

process, word-and-paradigm) abandon the notion of a morpheme and use the word as the basic unit (Bauer et al. 2013: 629). Word-formation rules or processes can be seen as independent entities belonging either to the lexicon (interfaced with grammar in various ways) or to grammar (and within grammar, either to a component of their own or to that shared with one or more levels of grammar). Alternatively, they can be seen as attached to morphemes or words, which will usually be stored in the lexicon. The end results, complex words, can be seen as either stored in the lexicon for reasons like ease of processing, not stored at all for reasons including economy and theoretical elegance, or partly stored depending on factors such as frequency and transparency (Plag 1999: 9–11; Brinton and Traugott 2005: 91).

To take an example of an approach belonging to the word-based item-and-process category, Construction Morphology does not see affixes as lexical items. Rather, word-formation patterns are regarded as "abstract schemas that generalize over sets of existing complex words with a systematic correlation between form and meaning" (Booij 2007: 34). The lexicon is a hierarchical structure consisting of both schemas and existing words, all of which are seen as constructions (form–content pairs) at different levels of abstraction. These constructions seem to be a special case of grammatical constructions, and thus the lexicon is subsumed under grammar, which is seen as a hierarchical inventory of constructions.

Hilpert (2013: 10) argues that the construction approach can be used to explain, e.g., the development of new word-formation patterns as a single process of constructional change, for instance "the grammaticalization of Old English *had* 'state, condition' into a suffix, the formation of lexical items with that suffix [-*hood*], and subsequent changes in the productivity of the suffix". If the entire process were to be called grammaticalisation, it would be difficult to explain decreases in the productivity of the suffix, as grammaticalisation is usually seen as a unidirectional process from less to more grammatical, and this is often seen to imply an increase in productivity. Furthermore, according to Hilpert (2013: 10), the lexicalisation of the lexical items would also be confusing in the grammaticalisation framework.

I agree with Hilpert (2013: 10, 114–115) that changes involving both increase and decrease in productivity seem to straddle the divide between grammaticalisation and lexicalisation. This is evident in Brinton and Traugott's (2005: 91–92, 109) account, which states that lexicalisation may result in "restricted" derivational affixes, while grammaticalisation may result in "default" affixes (whether inflectional or derivational), and that grammaticalisation implies increasing productivity, whereas lexicalisation implies decreasing productivity.

However, I do not think that the only way to study such changes is by a constructional approach, although such an approach is certainly a valid one.

Like Bauer et al. (2013: 628), I try to avoid committing to a single morphological theory in my work, the aims of which are primarily descriptive and methodological. Hence, the concepts introduced in Section 2.1 above are mostly based on relatively uncontroversial textbook definitions. The approaches to morphology that I find the most appealing are usage-based like Booij's (2010) Construction Morphology. However, the results of my studies are open to several theoretical interpretations and have also been cited by proponents of the generative tradition (Baeskow 2012). I, too, refer to work from a range of traditions, including structuralism (Marchand 1969), generativism (Aronoff 1976) and Natural Morphology (Dalton-Puffer 1996; Cowie 1999). The results I am the most inclined to trust are those based on robust corpus-linguistic methods and an adequate amount of data. In my own work, I therefore treat productivity as a quantitative notion, building upon theoretical and methodological contributions by Baayen (e.g., 1992).

2.3. Morphological productivity

Morphological productivity is a multi-faceted phenomenon; as Plag (2006: 547–549) shows, it is a derived notion instead of a theoretical primitive, but potentially useful in describing word-formation. Plag (2003: 44) defines productivity as the "property of an affix to be used to coin new complex words". This section discusses productivity as a qualitative, quantitative and diachronic notion, as well as structural, pragmatic and sociolinguistic factors conditioning and contributing to variation in productivity.

2.3.1. Productivity as a qualitative notion

Productivity can be conceived of as a qualitative, either–or notion: either an affix can be used to coin new words or it cannot. This view is advocated by, e.g., Bauer (1983: 99–100), who does not consider semi-productivity a useful construct. Plag (2006: 540), on the other hand, proposes three categories of morphological processes: those clearly unproductive, those clearly productive and those in between. I am not convinced of the usefulness of either of these views. It seems to me that an affix, or the process of forming words with it, can never be said to be clearly unproductive – there is always the possibility that somebody uses it to coin a new word (cf. Bauer et al. 2013: 641). This one-off use can be called

analogy instead of productivity, but where do we draw the line between the two; how many words must be coined for a process to be called productive?

This question is also posed by Plag (2006: 539–540), and it leads him to the three-way classification presented above, but that does not in my opinion really answer the question. Which would we classify as clearly unproductive and which as in-between? Besides, as Dalton-Puffer (1996: 222) points out, it is possible that analogy only differs from rule-based productivity in degree rather than in kind, so again there are no clear-cut boundaries (cf. Bauer 2001: 97 and the discussion in Section 2.1 above).

Furthermore, just like the distinction between clearly unproductive and in-between processes, the distinction between in-between and clearly productive ones is far from being straightforward. Again, how many new words must be coined for a process to be called clearly productive rather than in between; or are there some other criteria by which the classification can be made? Plag's exact definition of the in-between category is "those processes that are not easily classified as either productive or unproductive" (2006: 540) – I think most, if not all, processes would fall into this category, which would make the categorisation somewhat pointless.

Therefore, it seems to me that rather than asking *whether* a process is productive or unproductive or semi-productive, a better research question would be to ask *how* productive it is along some scale (or several), perhaps in comparison with another process or among different groups of people. This is, in fact, precisely what I aim to do in the present work. This scalar view of productivity will be discussed in the following section.

2.3.2. Productivity as a quantitative notion

Productivity can also be conceived of as a quantitative notion: an affix can be used to coin new words to some degree. Several ways of measuring this degree have been proposed in the literature. Baayen (1993) presents three measures, which he calls the category-conditioned degree of productivity (P), the hapax-conditioned degree of productivity (P^*) and the activation level (A). All of these are based on counting **tokens** (N) and **types** (V) of words belonging to a certain morphological category – for example, how many instances of *-ness* words and how many different *-ness* words a corpus contains, respectively. Types can also be used on their own as a measure of the **extent of use** or the realised productivity of an affix (Baayen 2009). Of special interest are the so-called hapax legomena or **hapaxes** (n_1), words that occur only once in the corpus, because these are seen to predict the number of new words.

According to Baayen and Lieber (1991: 810), a large proportion of the types of productive affixes are hapaxes, and the frequency distribution of the types is asymmetrical in general: there are more types that occur once than those that occur twice, more types that occur twice than those that occur three times, and so on. Overall, there are many types that occur only a few times in the corpus, and few types that occur many times. With less productive categories, the number of hapaxes is lower (there may be more **dis legomena**, types that occur twice, than hapaxes), and the frequency distribution is less skewed.

The **category-conditioned degree of productivity** P is defined as the ratio between the number of hapaxes with a given affix and the total number of tokens with that affix in the corpus: $P = n_1/N$. According to Baayen and Lieber (1991: 809–810), it expresses the probability of observing new types with the relevant affix when N tokens with the affix have been sampled. If the size of the corpus is increased, N will increase, and so will the number of types V, but at a different rate from N (Baayen and Lieber 1991: 811). This can be illustrated by drawing a graph with the values of V at different points of sampling on the y-axis and the values of N on the x-axis; in other words, V may be plotted as a function of N, $V(N)$. See Figure 2.1 for a schematic example using not affix types and tokens but all of the different words and the number of running words in a text.

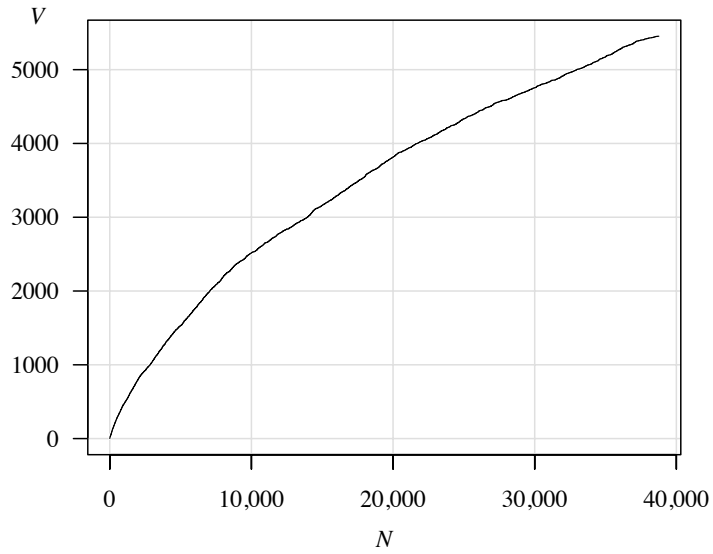

Figure 2.1. The growth curve of all types *V* as a function of all tokens *N* in the Project Gutenberg e-text of Joseph Conrad's *Heart of Darkness*, <http://www.gutenberg.org/2/1/219/> (see Chapter 9 below)

When only a few tokens have been observed, i.e., when *N* is small, new types will be found quite frequently, i.e., *V* will grow rapidly. As more and more types are found, the rate of growth will slow down. It is precisely this rate of growth that is expressed by the category-conditioned degree of productivity *P*. If the number of tokens observed is *M*, then *P*(*M*) is the slope of the tangent to the growth curve of *V* in the point (*M*, *V*(*M*)). A large value of *P*(*M*) indicates that there are many types yet to be sampled, which would suggest that the affix is productive. If, on the other hand, *P*(*M*) is small, the growth curve is about to even out, and the number of new types to be expected is small, which would suggest that the affix is unproductive (their term) or less productive (my term). (Baayen and Lieber 1991: 811–812.)

Baayen and Lieber (1991: 817) point out that once we have calculated the *P* for all of the tokens *N* in our sample, we know little about when the growth curve of *V* would flatten out if the sample size were increased; we cannot use *P* to predict the absolute number of types that would be found in a larger sample. In other words, *P* is dependent on the size of the corpus. Furthermore, because *P* is based on the number of tokens of a particular morphological category, it cannot

be directly compared with a *P* calculated from the number of tokens of another morphological category, unless the numbers of tokens are of a similar magnitude (Baayen 1993: 191). For example, we cannot compare the degrees of productivity calculated for *-ness* and *-ity* in the same corpus if there are many more tokens of *-ness* than of *-ity*.

Baayen's (1993) second measure of productivity is the **hapax-conditioned degree of productivity** P^*. It is defined as the ratio between the number of hapaxes with a given affix and the total number of all hapaxes in the corpus: $P^* = n_1/h$. This measure indicates how much the affix contributes to the overall vocabulary growth of the corpus (Baayen 1993: 193). According to Hay and Baayen (2003: 101), the hapax-conditioned degrees of productivity of different affixes within the same corpus can be compared by using n_1 counts alone, h being constant.

Baayen also has a third measure of productivity, the **activation level** A (1993: 195–196). It is defined as the number of tokens representing those types of a given affix whose frequency of occurrence is smaller than a threshold θ. The measure is motivated by the idea that people process complex words through two competing routes simultaneously: by **parsing** and by retrieving the whole word directly from their mental lexicon. Which route is faster depends on the frequency of the word: common words are readily available in the lexicon, while words rarer than the threshold value are processed by identifying and combining the affix and base. This process of parsing maintains the activation level of the affix; the level indicates how quickly the affix will be recognised and combined with the base.

Baayen (1993: 203) admits that choosing a suitable threshold value is problematic. He assumes that only words that are semantically transparent will maintain the activation level. The higher θ is, the more semantically opaque words it will include: words are usually the more opaque the more frequent they are, because frequent words have accumulated more meanings or their meaning may have changed. Therefore, the threshold is best kept fairly small. Hay (2001) improves on this measure by introducing the concept of relative frequency; she shows that it is not the frequency of the affixed word alone that affects parsability, but rather the relation between the frequencies of the affixed word and its base.

In addition to the semantic transparency mentioned above, phonological transparency has also been shown to affect productivity. According to Hay and Baayen (2003: 105), there is a wealth of evidence that English-speaking people use phonotactic probabilities when parsing words they hear. For example (2003:

105), as the /pf/ transition in *pipeful* is unlikely to occur within a monomorphemic word in English, people hearing this combination in running speech will immediately posit a morpheme boundary between *pipe* and *ful*, which will help them to decompose the complex word. These unlikely combinations facilitate the parsing route rather than the whole-word route of processing; therefore, complex words containing unlikely transitions at morpheme boundaries are more likely to be stored separately as affix and base rather than as a whole word in the mental lexicon; this in turn means that the affixes contained in complex words like these are more likely to be used productively (2003: 105–106). Hay and Baayen conclude that the mean probability of juncture created by the affix is correlated with productivity and could be used as an additional measure or predictor of productivity.

Out of these measures, three are commonly used and recommended by, e.g., Baayen (2009): the extent of use *V*, also known as **realised productivity**; the category-conditioned degree of productivity *P*, also known as **potential productivity**; and the hapax-conditioned degree of productivity *P**, also known as **expanding productivity**.

2.3.3. Productivity as a diachronic notion

Romaine (1985) is one of the first to explicitly consider the theoretical and methodological issues in productivity from a diachronic perspective. Even though a diachronic approach to productivity is in her view essential if we wish to understand the synchronic situation, Romaine (1985: 457–458) sees some problems with it. There is no access to informants' intuitions about possible words; furthermore, dictionary listings of actual words from earlier periods are even less comprehensive than listings of present-day words, and the first attestation dates in, e.g., the *Oxford English Dictionary* (OED) are unreliable.

Starting from the assumption that change stems from variation, Romaine (1985) hypothesises that competing patterns of word-formation establish themselves through social or stylistic specialisation – according to Romaine, they become markers. An important pragmatic factor affecting productivity is the demands of different genres and the development of new genres over time. Riddle (1985) argues that competing affixes may also undergo semantic specialisation proceeding through lexical diffusion. She criticises the sort of morphological experiments and analyses in which the possibility of a semantic distinction between competing affixes is not taken into account. A similar position is taken by Lloyd (2011).

Methodologically, historical studies have in general been less sophisticated than studies of Present-day English (PDE), partly because there has been too little data for proper statistical analysis. While psycholinguistic experimentation has been impossible for obvious reasons, scholars have developed complementary methods, such as counting hybrid formations that combine native and foreign elements (Romaine 1985; Dalton-Puffer 1992, 1996; Hlava 2005; Palmer 2009; Gardner 2013) or using first attestation dates from dictionaries (Barber 1976; Wermser 1976, as cited in Nevalainen 1999a; Anshen and Aronoff 1989; Lindsay and Aronoff 2013). As noted by Cowie and Dalton-Puffer (2002) and Hoffmann (2004), the coverage of the OED varies widely between different periods in terms of both the number of citations and text types, making diachronic comparisons unreliable. Here as in PDE studies, corpus linguistics is a welcome addition to the dictionary-based approach, as it provides information on actual usage (Březina 2005: 158–159). Cowie and Dalton-Puffer (2002) suggest that the main problem with historical corpora is their small size; however, we are now beginning to have access to immense databases such as *Early English Books Online* (EEBO), which may somewhat alleviate this issue.

The diachronic perspective highlights some of the problems with using type counts, or the extent of use *V*, as a measure of productivity (see Cowie and Dalton-Puffer 2002). Firstly, the existence of a large number of types may be due to aggregation through productivity in the past rather than current productivity (Anshen and Aronoff 1989). Secondly, in the case of etymologically foreign affixes, some words have been borrowed as a package including the affix, with no productivity involved at all in English (e.g., Marchand 1969: 313). Thirdly, an increase in the number of types between one time period and the next in a corpus does not straightforwardly mean an increase in productivity, because the types are not necessarily aggregated. Fourthly, comparing type counts obtained from (sub)corpora of varying sizes is difficult, as normalisation is not a legitimate option. Another problematic concept is that of "a point in time" (Cowie and Dalton-Puffer 2002: 420–421).

Cowie and Dalton-Puffer (2002: 431–432) also criticise hapax-based measures of productivity on the grounds that hapax legomena will almost never represent genuinely new formations in small corpora. As Baayen (2009: 905–906) points out, however, hapaxes do not have to be neologisms to be able to estimate vocabulary growth: they are a theoretically motivated tool, not an end *an sich*. The validity of hapax-based measures is discussed further in Chapter 9.

Using the OED quotations database as a corpus, Hilpert (2013: 127–133) discusses the applicability of the productivity measures *V*, *P* and *P** to diachronic

dictionary data. He notes that the problem of unequal period sizes makes the measures *V* and *P* unusable, and equalising the sizes of the time periods by taking a random sample from each (as suggested by Gaeta and Ricca 2006) would result in too little data in the case of the suffix *-ment*. Hilpert assumes that the measure *P**, which is the ratio between affix hapaxes and all hapaxes in the (sub)corpus, is not dependent on corpus size in a similar manner, and decides to use it as the basis for further statistical analysis. Nevertheless, as gaining access to all hapaxes in the OED via the online interface would be impossible, he modifies the measure by only counting the hapaxes in the quotations containing a *-ment* word. This means that the results are only comparable internally, not with other affixes. On the other hand, as the *P** measure estimates the contribution of the affix to overall vocabulary growth, it could be argued that other constructions are taken into account indirectly. Hilpert (2013: 133–134) goes on to use *P** in a technique called Variability-based Neighbour Clustering to discover an optimal periodisation of the OED for the purpose of studying constructional change in *-ment*.

Further diachronic research on productivity is surveyed in Section 2.4 on *-ness* and *-ity*. The next section concentrates on synchronic variation from which change is assumed to arise, taking into account both intra- and extralinguistic factors.

2.3.4. Variation in productivity

While many scholars only discuss 'restrictions' or 'constraints' on productivity, Cowie argues that "restrictive factors must be distinguished from factors which are the primary motivators of lexical innovation" (1999: 83) and that the latter are "chiefly extralinguistic in nature" (1999: 6–7). As structural and pragmatic factors contributing to variation in productivity have been reasonably well studied, Sections 2.3.4.1 and 2.3.4.2 are mainly based on textbook and handbook material that provide an excellent overview of the factors. No such material is yet available for sociolinguistic factors, which have only recently started to attract scholars' attention (but see Körtvélyessy and Štekauer forthcoming), so Section 2.3.4.3 surveys the individual studies that have been conducted so far.

2.3.4.1. Structural factors

According to Plag (2006: 550–551), structural factors influencing productivity can be divided into phonological, morphological, syntactic and semantic constraints. Furthermore, they can be either general or process-specific; the latter may relate to what the base or the derived word must be like. As Plag (1999: 43–

44) states, however, the boundaries between these divisions can be fuzzy. Let us first have a look at examples of process-specific constraints (Table 2.1 below). Here the first three constraints apply to the base and the last one to the derived word.

Table 2.1. Process-specific structural constraints on productivity (Plag 2006: 551)

Type	Example	Constraint
Phonological	Suffixation of verbal *-en* (as in *blacken*)	Only attaches to base-final obstruents, does not take bases that have more than one syllable
Morphological	Suffix combination *-ize-ation*	Words ending in the suffix *-ize* can be turned into a noun only by adding *-ation*
Syntactic	Adjectival suffix *-able*	Normally attaches to verbs
Semantic	Suffix *-ee* (as in *employee*)	Derivatives with the suffix must denote sentient entities

As for general structural constraints, Plag (1999: 45) lists ten of these but swiftly dismisses most of them. Among the more tenable ones is the **unitary output hypothesis**, which states that derivatives from a certain word-formation process form a group uniquely distinguishable from others through its phonological, syntactic and semantic properties (Plag 1999: 49). While Bauer (2001: 127–128) points out that semantic unity or disunity is in the eye of the analyst, he admits that the hypothesis is relatively uncontroversial.

Another general constraint is **blocking**, which is defined by Aronoff (1976: 43) as "the nonoccurrence of one form due to the simple existence of another". Van Marle (1986: 607) distinguishes between two special cases, which Rainer (1988: 159) calls token-blocking and type-blocking. According to Rainer (1988: 161), **token-blocking** occurs when the creation of a morphologically complex word, such as *stealer*, is blocked by an existing synonymous word, such as *thief*. Rainer shows that it does not matter whether the blocking word is idiosyncratic (like *thief*) or regularly derived, as long as it is stored in the lexicon (1988: 164–167).

According to Rainer (1988: 161–164), token-blocking may only occur under the three conditions of synonymy, productivity and frequency. Firstly, the blocked word and the blocking word must be truly synonymous; secondly, the blocked word must be a potential word in the sense that there is a productive

morphological process by which it could be formed; and thirdly, the blocking word must be frequent enough to be retrieved from the lexicon faster than the blocked word can be formed.

Van Marle (1986: 613–617) notes that some blocking words radiate a stronger "blocking force" than others; Rainer (1988: 163–173) shows that the strength of the force depends on the frequency of the blocking word. The less frequent the stored word is, the greater the likelihood that the speaker will fail to activate it, which according to Plag (2006: 552) explains the occasional failure of blocking and the occurrence of synonymous doublets. However, Rainer (1988: 164) remarks that the blocking force is resisted by another force dependent on the productivity of the morphological process that would produce the potential word; it is the interplay of these two forces that actually determines whether blocking succeeds or fails.

According to Plag (2003: 66), **type-blocking** "has been said to occur when a certain affix blocks the application of another affix". Van Marle's (1986: 608) domain hypothesis suggests that affixes can be divided into two groups: special cases, which can only be applied within a restricted domain subject to specific constraints, and general cases, whose domains are unrestricted except for the paradigmatic limitation that they do not include bases belonging to the domains of rival special cases. For example, *-ness* suffixation could be seen as a general case that is blocked by the special case of *-ity* suffixation (Plag 2006: 552).

Plag (2006: 552–553) shows, however, that there are at least three problems with this kind of analysis: Firstly, even though one of Rainer's (1988: 173) preconditions for type-blocking is synonymy, *-ness* and *-ity* are not always synonymous, as shown by Riddle (1985). Secondly, there are plenty of attested doublets, which means that the domains do not completely exclude one another. Thirdly, it is unclear how putative cases of type-blocking can be distinguished from token-blocking – if some form in *-ness* is avoided, how can we rule out the possibility that it is because the equivalent form in *-ity* exists in the lexicon? The first two problems may not apply to all cases, but I think the third problem is a crucial one. Plag (2003: 67–68) in fact suggests that we should reject the notion of type-blocking altogether, as we can only verify the existence of token-blocking. I agree; however, I think this should not be taken as an indication that affix rivalry does not exist but simply that it cannot be adequately described with the concept of type-blocking as defined above.

2.3.4.2. Pragmatic factors

One productivity constraint that I think could be classified as pragmatic is the usefulness of the potential word to members of the speech community. Kastovsky (1986: 594–595) names two functions of word-formation: labelling and syntactic recategorisation. The former is used to refer to a new concept or object, while the latter replaces a phrase or a clause with a single complex word in order to condense information, create stylistic variation or facilitate text cohesion (Plag 2003: 59–60). The two functions are exemplified in (2.3) and (2.4), respectively.

(2.3) The Time Patrol also had to *unmurder* Capistrano's great-grandmother, *unmarry* him from the pasha's daughter in 1600, and *uncreate* those three kids he had fathered. (Kastovsky 1986: 594)

(2.4) If that's not *civil*, *civilize* it and tell me. (Kastovsky 1986: 595)

Plag (2003: 59–60) adds a third function, namely that of expressing an attitude, as in (2.5). This function was also discovered by Renouf and Baayen (1998), who found that *mock-* was used to indicate irony in British journalism.

(2.5) Come here, *sweetie*, let me kiss you. / Did you bring your wonderful *doggie*, my darling? (Plag 2003: 59)

According to Plag (2006: 550), one of the most important pragmatic factors affecting productivity is fashion. As noted by Renouf (2007: 87–88), there is a kind of ebb and flow in what is in vogue and what is not. For example, extralinguistic developments can be seen to have influenced the extent of use of the affixes *mega-*, *giga-*, *mini-* and *-nik* (Plag 2006: 550). The second pragmatic constraint mentioned by Plag is that new words must "denote something nameable" (2006: 550). By this he means that the new concept cannot be overly complex – typical derivative affixes only add a very simple and general meaning to that denoted by the base (e.g., adjectival *un-* 'not X').

Plag (2003: 61) warns that we should not automatically assume a pragmatic reason for why some new formation is impossible – there may well be structural constraints involved, and the existence of these should in fact be checked before entering into any usage-based speculations.

2.3.4.3. Sociolinguistic factors

As noted in Section 2.3.3, Romaine (1985) hypothesises that social and stylistic factors influence productivity. While stylistic factors have been studied since the 1990s (Baayen 1994; Baayen and Neijt 1997; Plag et al. 1999; Cowie 1999), chiefly in terms of genre or register (see 3.3.3 below), the study of sociolinguistic variation in morphological productivity has only begun to gain ground in the past decade or so, and much remains to be done. In an early experimental study, Romaine (1983) analyses sociolinguistic variation in acceptability judgments of *-ness* and *-ity* formations, finding a considerable amount of individual variability as well as some patterns based on age and gender. However, according to her it is unclear whether this variability is related to productivity or to the metalinguistic ability to make judgments of this kind. There is thus a need for other types of evidence.

Taking an onomasiological approach, Štekauer et al. (2005) study sociolinguistic variation in speakers' preferred naming strategies in forming new words. They operationalise the productivity of an individual naming strategy as the share of the strategy out of all complex words within a given cognitive category, at several levels. For instance, at the level of word-formation rules, their experiment shows that the most productive rule belonging to the Agentive category in English is Object–Action–Agent (N V *-er*), which constitutes 20.92% of native speakers' formations. They find sociolinguistic variation in productivity such that older and more highly educated people prefer more explicit naming strategies (such as the three-constituent N V *-er*), as do non-native speakers, especially those whose mother tongue is agglutinating. By contrast, more creative and economical naming strategies are preferred by younger and less educated speakers, who also more readily accept formations as grammatical (cf. Romaine 1983). The onomasiological approach is continued by, e.g., Körtvélyessy (2009).

Schröder (2008: 190–208) analyses sociolinguistic variation in acceptability ratings of verbal prefixation in English. She finds no clear gender differences, and only a slight tendency for younger speakers to be more tolerant of new or rare formations (cf. Romaine 1983; Štekauer et al. 2005). Contrary to Štekauer et al. (2005), she discovers that highly educated people also exhibit slightly more tolerance towards new formations.

In an innovative MA thesis, Březina (2005) combines derivational morphology with historical sociolinguistics by investigating gender variation in the use of the prefixes *in-* and *un-* in the *Corpus of Early English Correspondence Sampler*. Apparently basing his conclusion on token counts only, Březina (2005: 147) finds that "women's use of the *un-* and *in-* forms is more progressive than men's use".

An examination of the frequency of each type with men and women, however, reveals that "women repeatedly use a more limited number of *in-* forms" and that 78% of women's total use of *in-* comes from their 20 most frequently used types, while the men's corresponding figure is 63% (Březina 2005: 151).

Palmer (2009: 274–280), too, uses normalised token frequencies in his analysis of gender variation in the use of nominal suffixes in the *Corpus of Early English Correspondence*. Gardner (2013: Chapter 5) discovers regional variation in the productivity of nominal suffixes in corpora of Middle English by comparing the frequencies of new types, i.e., the first occurrences of those types in the corpora that have not been documented in Old English. The studies by these two authors will be discussed in more detail in the next section.

More complex methods are used by Keune et al. (2006) in their synchronic study of the *Corpus of Spoken Dutch*. They analyse the effect of country, gender, education and age on the productivity of 72 affixes in terms of hapax frequency. Comparing three different statistical models which they fit to their data, they find that the generalised linear model performs best. While I do not think that this model takes into account the non-linear dependence of hapax frequency on corpus size, it does enable Keune et al. to gain results that are linguistically interesting. For instance, they find that the highest affixal productivity is generally exhibited by highly educated older men (cf. Štekauer et al. 2005).

Using similar methods, Keune (2012: Chapter 4) observes significant sociolinguistic and register variation in productivity in spoken and written Dutch, discovering parallel patterns of variation in both derivational productivity (estimated by affix hapaxes) and lexical productivity (all hapaxes). Again, highly educated older men emerge as the most productive group. Individual affixes may deviate from this pattern, however – for instance, diminutive suffixes are used more productively by women than by men, and the prefix *super-* is preferred by the youngest age group. Furthermore, the effects of age and education seem to disappear in private, spontaneous speech representing a more involved style.

Lupica Spagnolo (2013) compares the productivity of the affixes *-ung*, *-keit*/*-igkeit*/*-heit*, *-bar* and *ver-* in a native and non-native corpus of German literary texts, finding no significant differences between native and non-native authors. She uses Baayen's (2009) productivity measures in combination with Gaeta and Ricca's (2006) variable-corpus approach, i.e., taking samples of equal size in terms of token frequency, to be able to compare measurements across subcorpora (note, however, that this approach is not always applicable; see 2.3.3 above). To determine the statistical significance of the observed differences, she uses the test

of equal or given proportions in R (R Core Team 2014), which is similar to the χ^2 test (see 5.1.1 below).

In summary, social categories that have turned out to be relevant in the context of morphological productivity include gender, age, education and region. Methodology-wise, no satisfactory solution has been provided to the problem of comparing productivity measurements obtained from sociolinguistically defined subcorpora of varying sizes.

2.4. Previous research on *-ness* and *-ity*

Previous studies of our pair of suffixes have produced varied results depending on the kind of material and methods employed. This section provides an overview of previous research through a diachronic survey of the suffixes from Old English to Present-day English, with change in semantics and adjectival base types discussed at the end of the section. With historical studies, it is useful to bear in mind that the results are often based on relatively small data sets, which are used in different ways by different scholars. This may affect the reliability of the results and may also explain why the results from different studies are not always in agreement.

The main results from historical studies concerning the productivity of *-ness* and *-ity* throughout the centuries can be summarised as follows. The native suffix *-ness* dates from before the Old English period; originally, it was mainly used with verbs, but since Old English it has most frequently been based on adjectives (Romaine 1985; Riddle 1985; Dalton-Puffer 1996: 81–85). Noun bases are infrequent but occur in all periods of English, and other parts of speech excepting verbs seem to have become increasingly possible as bases (OED, s.v. *-ness*, suffix; Marchand 1969: 334–336), which may indicate an increase in productivity, or possibly a change in the semantics of the suffix (Riddle 1985; cf. Romaine 1983; Baayen and Renouf 1996: 84–85). Competition between *-ness* and other native processes, such as *-dom*, *-hood* and *-ship*, was largely resolved in favour of *-ness*, which became the most productive of the deadjectival suffixes forming abstract nouns by the end of the Old English period (Romaine 1985).

In Middle English, *-ity* entered the language through loanwords from French and later also through calques on Latin; while it did not reach the level of productivity of *-ness*, it began to compete with *-ness* as a more learned and prestigious deadjectival suffix meaning approximately the same thing (Marchand 1969: 312–313, 334–335; Barber 1976; Riddle 1985; Dalton-Puffer 1996: 120, 128). The productivity of *-ness* also increased during this time, mostly in the new literate genres (Romaine 1985).

Gardner (2013) sheds additional light on variation and change in the productivity of *-ness* and *-ity* in Middle English using four multi-genre corpora and several methods, including the one presented in Chapter 6 below. She finds that both deadjectival *-ness* and *-ity* reached their highest levels of productivity in the second half of the 14th century, decreasing towards the end of the Late Middle English period (Gardner 2013: 79–82, 108–111). In terms of regional variation, the East Midlands seem to play an important role in new *-ness* formations, with the focus moving to the West Midlands in Late Middle English, whereas the opposite is true for *-ity* (Gardner 2013: 120–122, 140–142). For *-ness*, regional variation is more important than genre variation, although religious, scientific and literary texts seem to dominate, but genre is clearly important in the case of *-ity*, which is especially productive in sermons, documents, literary texts and translations (Gardner 2013: Chapter 6).

Examining token frequencies in 15th- and 16th-century correspondence, Palmer (2009: 277–280) finds no significant gender differences in the use of *-ness*. Women tended to use more *-ity* tokens than men except for the first half of the 16th century, when men began to use highly learned derivatives in *-ity*, *-cion* and *-ment*. The 16th century also saw *-ity* becoming as productive as *-ness* in terms of type frequency, while the parsability, or perceived productivity, of both suffixes increased but with *-ness* clearly in the lead (Palmer 2009: 281–282, 298). Both *-ness* and *-ity* showed more parsability in medical texts than in correspondence (Palmer 2009: 306). According to Nevalainen (1999a: 357–358, 398), by Early Modern English both *-ness* and *-ity* were very productive, as new words were needed to facilitate the use of English in an ever-widening variety of functions.

Based on the OED, Lindsay and Aronoff (2013) find that the productivity of *-ity* undergoes a rapid increase during the 17th century and has continued to rise ever since. By contrast, analysing the proportion of new types in *-ness* and *-ity* in the ARCHER corpus from 1650 to 1990 using the first two periods of the *Helsinki Corpus* as the base lexicon, Cowie (1999: 193–195) finds that the proportions fluctuate with no continuous increase or decrease for either suffix. While *-ness* is more productive than *-ity* in 1650–1700 and 1800–1950, otherwise the suffixes are more or less equal. However, there is genre variation in the productivity of *-ness* and *-ity* such that *-ness* is preferred in sermons, fiction and correspondence, whereas *-ity* belongs more to the domain of scientific and medical writing (Cowie 1999: 248, 224). Cowie (1999: 224) tentatively connects *-ness* with Biber and Finegan's (1997) involved style and *-ity* with an

informational style, although sermons are not classified as involved by Biber and Finegan.

Most scholars seem to agree that *-ness* is more productive than *-ity* in Present-day English (e.g., Baayen 1993).[1] This varies by the type of base, however: *-ity* is more productive than *-ness* with bases in *-al*, *-able*, *-ible*, *-ic* and *-ile*, among others, while *-ness* outnumbers *-ity* with bases such as *-ive*, *-ous* and, of course, native bases (Marchand 1969: 314, 334–335; Aronoff 1976: 36; Aronoff and Schvaneveldt 1978; Anshen and Aronoff 1989; Baayen and Lieber 1991; Baayen and Renouf 1996). Plag et al. (1999) find that the productivity of both suffixes also varies by register: both are more productive in written than spoken language in general, and least productive of all in everyday conversations. Nevertheless, *-ness* is seen as the default suffix for forming de-adjectival abstract nouns (or, more inclusively, abstract nouns from non-verbal categories) by, e.g., Aronoff and Anshen (1998) and Bauer et al. (2013: 246).

The results from PDE studies also indicate that the overall lower productivity of *-ity* is due to factors at several levels of language. The phonology of *-ity* is more complex than that of *-ness* (Aronoff 1976: 38–43), and its junctural phonotactics is not sufficiently different from morpheme-internal transitions to facilitate parsing (Hay and Baayen 2003); morphologically, there are fewer bases to which it can attach, as native bases are usually out of the question (Marchand 1969: 312; Baayen and Lieber 1991; Hay and Plag 2004); semantically, its meaning is less coherent than that of *-ness* (Aronoff 1976: 38–43; Aronoff and Anshen 1998); and therefore, it is pragmatically better suited to the labelling function, which is not needed as often as recategorisation, at least not in written language (Plag et al. 1999; cf. Baayen and Neijt 1997). The functional/semantic differentiation between *-ity* and *-ness* is a process continuing to this day: while *-ity* is mainly used for labelling abstract or concrete entities, *-ness* is used more for syntactic recategorisation with the meaning 'embodied attribute or trait' (Riddle 1985; Plag et al. 1999; Baayen and Neijt 1997; Aronoff and Anshen 1998).

[1] According to Aronoff and Anshen (1998: 244–245), *-ity* seems to be more productive than *-ness* in 20th-century English, which goes against the results obtained by Baayen and his associates. As Aronoff and Anshen themselves admit, however, this result could be skewed: it is based on entries in the OED, and it may well be that *-ity* words are more likely to be listed in the dictionary simply because they are more unusual and thus more memorable than *-ness* words. Furthermore, as Baayen and Renouf (1996: 69) point out, it would be commercially unappealing for dictionary-makers to print all of the productively formed *-ness* words whose meaning is completely transparent to everybody anyway.

The semantic side of this equation, however, is challenged by Cowie (1999: 260–264), who argues, firstly, that Riddle (1985) is wrong in her claim that there was an influx of *-ity* words borrowed in the entity sense in Middle English, which would have started the differentiation. Secondly, Cowie shows that *-ity* words have been coined in the attributive meaning even in the 20th century, and while *-ness* words are difficult to use as labels, she argues that this is due to pragmatic rather than semantic reasons, namely "to the learned connotations of *-ity* and its use in specialized terminology". Thirdly, Cowie claims that it is impossible to consistently distinguish between the attributive and abstract entity senses other than by determining whether or not the nominalisation appears as the object of a genitive ("Catherine's boldness startled the elders" vs. "Boldness was frowned upon in the small community"), which according to her is not a change in semantic category. Cowie prefers Romaine's (1985) depiction of semantic change in *-ity* formations, which goes from abstract to concrete, as in the two senses of *curiosity*; this can also happen to *-ness* formations, but much more rarely.

Baeskow (2012), too, notes that the difference between the attributive and abstract entity senses manifests itself in syntactic terms, but she argues that this is indeed a semantic distinction that could also be called specific vs. generic. She agrees with Riddle (1985) that *-ness* is associated more with specificity and *-ity* more with genericity in Present-day English (see also Adamson 1989: 214). Contrary to Riddle's hypothesis of change, Palmer (2009: 286) observes a trend in 15th- and 16th-century letters such that *-ity* seems to have moved from abstract to attributive (my interpretation): while new 15th-century types mostly represent "abstract descriptive states", in the 16th century they become more about "aspects of human disposition", which may have been used more often in an attributive sense.

As for change in the productivity of *-ness* and *-ity* among different classes of adjectival bases, *-ness* has been commonly attached to adjectives of French origin since 1300, although native bases were still preferred in at least Early Modern English; by contrast, hybrid formations in *-ity* seem to always have been rare (Marchand 1969: 335, 312; Romaine 1985; Dalton-Puffer 1996: 81–85, 220–222; Nevalainen 1999a: 398). While words in *-ableness* seem to have appeared earlier than those in *-ability*, the association of *-ity* with *-able* may have arisen already in Middle English through *-able*/*-ability* word pairs borrowed from Latin; nowadays, *-ability* is by far the more common choice, provided that the semantics of the *-able* word matches that of *-ity* (Marchand 1969: 313–314; Dalton-Puffer 1996: 107; Riddle 1985). Similarly, *-ibility* has become more and more common

than *-ibleness*, whereas *-iveness* seems to have been consistently somewhat favoured over *-ivity* (Anshen and Aronoff 1989). Other bases with which *-ity* became productive from the 15th century onwards include *-ic*, *-al* and *-ar* (Nevalainen 1999a: 398). Lindsay (2012) argues that the Latinate bases are how *-ity* was able to reach productivity, as it found a niche in an emergent system.

3. Historical sociolinguistics

In addition to word-formation, the second major area of study with which this dissertation concerns itself is historical sociolinguistics. This chapter introduces the concept (Section 3.1), discusses the notion of the linguistic variable in the context of derivation (Section 3.2) and describes the social categories used in the studies (Section 3.3).

3.1. Introduction

In addition to contemporary studies, sociolinguistics has in the past three decades begun to be applied to historical material. According to Nevalainen and Raumolin-Brunberg (2003: 2), the first systematic attempt at this was made by Romaine (1982). Nevalainen and Raumolin-Brunberg themselves are pioneers in this field, which is now called **historical sociolinguistics** (Nevalainen and Raumolin-Brunberg 2012: 22–23). Besides their mother disciplines of sociolinguistics and historical linguistics, historical sociolinguists draw on social history to ensure the social and historical validity of their work (Nevalainen and Raumolin-Brunberg 2003: 8–11; see Nevalainen and Raumolin-Brunberg 2012: 27 for a more complete picture of related fields).

3.2. Constructing the linguistic variable

In variationist sociolinguistics, social categories are studied in relation to the **linguistic variable**, which Milroy and Gordon (2003: 88) define as a linguistic item with variant realisations that refer to the same thing but covary with different items or social categories. According to Milroy and Gordon (2003: 88), the use of a variant can be described quantitatively, in terms of percentages, rather than as an either/or situation. The methodology for investigating such variation was largely developed by Labov (1978 [1972], etc.) and was originally applied to phonological variables. The extension of the concept to morphology, lexis and syntax has given rise to the requirement of referential sameness, which is not always easy to achieve and is thus sometimes substituted with functional equivalence in discourse (Cowie 1999: 184–185; Tagliamonte and D'Arcy 2009: 72–74).

In my research, I could regard *-ness* and *-ity* as variants between which speakers can choose when they wish to form an abstract noun meaning something like 'the property of being X'. This is, however, not an unproblematic point of view to take. As discussed by Cowie (1999: 186–188), even if we assume

referential sameness between *-ness* and *-ity*, each lexical item formed using either suffix will be in a paradigmatic relationship with other items, which may range from simplex words to syntactic constructions. Even the paradigm of synonymous nominal suffixes is not limited to *-ness* and *-ity*: in various phases of the history of English, it has included suffixes like *-th*, *-dom*, *-hood*, *-ship* (cf. 2.4 above) and, on a very limited number of bases, *-acy* and *-ancy*/*-ency* (Cowie 1999: 206–208). Furthermore, the number of bases actually shared by *-ness* and *-ity* turns out to be surprisingly small in many corpora (e.g., Cowie 1999: 198; Säily 2008: 71–72). It is therefore unclear what exactly should be included in the linguistic variable.

Aronoff and Gaeta (2003: 5) suggest that variation in the productivity of morphological processes can – and perhaps should – be examined separately for each process, and this is the approach taken in the present work. It is nevertheless illuminating to compare the kinds of variation observed in *-ness* on the one hand and *-ity* on the other. Methodologically, the absence of a linguistic variable means that productivity cannot be expressed in terms of, e.g., the proportion of *-ness* types out of the combined frequencies of *-ness* and *-ity* types, which again brings us to the difficulty of comparing type frequencies across subcorpora as a central issue addressed in this dissertation.

3.3. Social categories

The focus of this dissertation is on macro-level social categories (Labov (1978 [1972]: 183) within the framework of quantitative sociolinguistics. This section discusses the social categories employed in the present work – gender (3.3.1), social rank (3.3.2), register and genre (3.3.3) – and how they were realised in the sociohistorical contexts of the material used, chiefly those of England in the 17th, 18th and late 20th centuries. The categories were chosen based on previous research and on what it was possible to study in the material (see Chapter 4 below) in terms of metadata and coverage.

3.3.1. Gender

A social category that has proved to be a strong factor in language variation and change is **gender**. Like Nevalainen and Raumolin-Brunberg (2003: 110), I use the term gender instead of sex, because gender is a social construct, the characteristics of which can change over time. According to Milroy and Gordon (2003: 103), one generalisation to be made from previous work on present-day sociolinguistics is that women seem to prefer supralocal forms, i.e., ones that are fairly widely distributed, whereas men prefer local forms, which are often stig-

matised. Thus, women are often the leaders of supralocal language change, also known as supralocalisation, in which a linguistic feature spreads from one region to neighbouring areas (Nevalainen and Raumolin-Brunberg 2003: 112).

The category of gender is a pertinent one in 17th-century society as well. Women had an inferior status in comparison with men, and their rank came from the rank of their father or husband. The husband was the head of the household, and wife-beating was allowed by law, though frowned upon by people. The letters of gentlewomen to their husbands in this period show an anxiety to please (Wrightson 1993: 94), and women were expected to speak modestly in mixed society; in all-female contexts they could speak more freely (Mendelson and Crawford 1998: 212–213). Lower down on the social scale women could be more assertive in their speech (Wrightson 1993: 96), and rhetoric (including scolding, gossip, storytelling and folklore) was indeed one form of empowerment for women (Mendelson and Crawford 1998: 215–218).

Whereas men could freely move in both public and private spheres, women were mostly confined to the private sphere, with little opportunity for higher education or participation in the running of the society. According to Mendelson and Crawford, the situation changed somewhat during the Civil War, but after the Restoration women were forced to retreat from the public sphere at least formally (1998: 401–402, 419). As for employment, in addition to taking care of household chores in their own home, women could work as servants (young ladies of the gentry could be in the service of a noble kinswoman or the queen), and the poorer sort assisted their husbands in the shop or even worked in the field (Laslett 1965: 2–3, 11–12). Mendelson and Crawford say that women were on the whole less mobile than men (1998: 301), which could mean that in this period it was men rather than women who were the leaders of supralocal language change. On the other hand, women would often move when they got married, so they were not strangers to mobility.

Around the year 1600, only about a third of English males was literate in the sense of being able to both read and write (reading was taught before writing, and they were seen as two different skills). Women's level of literacy was much lower throughout the Early Modern English period. In the 16th century, most of the male gentry (including nobility) had been educated by domestic tutors, but in the 17th century this was replaced by formal education in grammar schools teaching humanist rhetoric and religious knowledge. Similarly, at the age when they would have formerly gone into service in noble households, it was now fashionable to send young gentlemen to universities and the Inns of Court. Girls, of course, could not go to either grammar school or university; gentry females may well

have been taught by domestic tutors, but the rest would only have learned to read and write in petty school, if that. There was thus a hierarchy of education reflecting rank, wealth and gender. (Wrightson 1993: 188–193.)

In the 18th century, the situation was broadly speaking similar. As noted in Chapter 10, education was more widely available but still stratified, universities being reserved for men of the "better sort" (Cannadine 2000 [1998]: 47–48). Most women of the "better sort" were literate, and some had received a high-level education at home, although this was not necessarily encouraged by society at large (Tieken-Boon van Ostade 2010). Notably, this period saw the rise of the group of educated and intellectual women known as the Bluestockings. McIntosh (2008: 231) goes so far as to say that 18th-century British culture was "feminized": there were more female authors than ever before, and the feminine virtues of politeness and sensibility were required of anyone aspiring to be counted among the upper classes.

In the late 20th century, the opportunities available to men and women were more equal, but the fields in which they were educated and the jobs they had were still highly gendered, and employment rates were higher for men than for women (Office for National Statistics 2013). Women were still granted less status and power than men, their sexuality remained more strictly controlled than that of men, and men were still expected to behave in a 'manly' manner like their peers, although the definition of manly behaviour may have changed over the centuries (Nevalainen 1999b: 511–513; James 1996; Milroy 1999; cf. Fletcher 1996; Foyster 1999).

3.3.2. Social class or rank

According to Milroy and Gordon (2003: 95), the category of **social class** comes from sociology, where it is used in two different models. The first model goes back to functionalist sociology and describes social classes as a flexible continuum of shared values and *consensus*, while the second model developed by both Marx and Weber treats class divisions as discrete and based on *conflict*. The consensus model sees occupations as the main way of distinguishing between different classes, whereas the conflict model gets its divisions from people's different relations to the market.

Milroy and Gordon (2003: 96–97) argue that Labovian sociolinguistics, with its peaceful and harmonious concept of speech community, overwhelmingly follows the consensus model. There are nevertheless some sociolinguists who question it, pointing out that the very fact that there are so many vigorous non-standard vernacular communities could be interpreted as evidence of conflict.

Indeed, if there were no conflict, where would language change come from? However, Milroy and Gordon note that both models are potentially useful in sociolinguistics and that different kinds of data require different approaches.

In the 17th century, we talk about social **rank** rather than class. As Laslett (1965: 22) explains, the class system had not yet arisen – if class is defined as "a number of people banded together in the exercise of a collective power, political and economic", in this pre-industrial society there was only one class, and most people would not have belonged to it. People did have different status levels, however, and for those levels I (following Nevalainen and Raumolin-Brunberg 2003: 33) choose to use the contemporary term rank.

There are many ways of dividing 17th-century society into ranks. A rough dichotomy would be gentry vs. non-gentry (Laslett 1965: 26); a tripartite division into the better sort, the middling sort and the poorest sort has also been suggested by Nevalainen and Raumolin-Brunberg (2003: 136, Model 4).[1] In Nevalainen and Raumolin-Brunberg's (2003: 136) Model 3, these contemporary labels are exchanged for the more neutral upper, middle and lower ranks, with the additional category of social aspirers above the middle ranks. Social aspirers were middle- or lower-ranking people who advanced to the upper ranks – for instance, merchants who became gentlemen or members of the upper clergy – and who would have wished to show their learning and gentility even in their language use.

Nevalainen and Raumolin-Brunberg's (2003: 136) Model 2 is an even more fine-grained division, with royalty at the very top, followed by nobility, gentry, clergy and social aspirers; next professionals (e.g., army officers, lawyers, medical doctors and teachers) and merchants; and, finally, other non-gentry (such as yeomen, husbandmen, craftsmen, labourers, cottagers and paupers). In Nevalainen and Raumolin-Brunberg's (2003: 136) Model 1, their most fine-grained model, the basic distinctions are the same as in Model 2, but the gentry is further subdivided into upper gentry (consisting of knights and baronets) and lower gentry (including esquires and gentlemen); furthermore, both upper and lower gentry are divided into a non-professional and a professional section, the latter of whose members held high government offices (2003: 137). Clergy, too, is divided into upper (bishops) and lower (the rest).

According to Wrightson (1993: 27–30), the rural gentry was as a rank preferred to the urban merchants and professionals; nevertheless, they were closely related, as merchants and professionals were often the younger sons of

[1] The better sort consists of royalty, nobility, gentry and clergy, while the middling sort are professionals and merchants, and the poorest sort other non-gentry (see Models 1–2 below).

gentry, and a successful merchant or professional could acquire land and retire to the country, thus becoming a part of the gentry. The ownership of land and freedom from manual labour were the crucial criteria in deciding who was a gentleman (1993: 25). There were other avenues open to urban aspirers besides the rank of a country gentleman, however: professionals could advance to powerful governmental positions, and merchants could be active in guilds and the government of the city.

In 18th-century England, the crucial division in society was still between gentry and non-gentry (Hay and Rogers 1997: 18–24). However, as the number of tradesmen and manufacturers increased dramatically during the period (Fitzmaurice 2012), the line became more blurred than before, so that landownership or even freedom from manual labour was no longer essential, and wealthy merchants and sons of great manufacturers could be called gentlemen. According to Vartiainen et al. (2013: 235), "[t]his blurring of lines – gentlewomen somewhat more on a par with gentlemen in terms of education and literary influence, and gentry a more inclusive category than before – may have had an effect on language use as well".

According to Cannadine (2000 [1998]: 171), social class was still alive and well in the late 20th century, especially in people's minds and discussions. Statistics show that class also correlated with tangible issues like health (Office for National Statistics 2004). A common way of classifying people into social groups was based on the occupation of the head of the household (National Readership Survey 2014): upper middle class (A; higher managerial, administrative and professional), middle class (B; intermediate managerial, administrative and professional), lower middle class (C1; supervisory, clerical and junior managerial, administrative and professional), skilled working class (C2; skilled manual workers), working class (D; semi-skilled and unskilled manual workers) and non-working (E; state pensioners, casual and lowest grade workers, unemployed with state benefits only). This is also the classification used in the *British National Corpus* (see Section 4.3 below).

3.3.3. Register and genre

Following Nevalainen and Raumolin-Brunberg (2003: 190), my concept of **register** comes from Halliday (1978): registers are varieties based on use. For the purposes of linguistic analysis, registers consist of three situational dimensions, namely **field** (type of activity and topic), **tenor** (addressee and other participant relations) and **mode** (channel of communication) of discourse. In Chapter 11, I apply this concept to 18th-century correspondence, where the mode is written and

the field is letter-writing, each letter covering a variety of topics. While most of the previous studies of register variation in productivity and in corpus linguistics in general have focussed on mode (e.g., spoken vs. written) and field of discourse, the kind of register variation I am interested in concerns tenor, i.e., the relationship between the sender and recipient of the letter. My hypothesis is that derivational productivity may vary in correspondence depending on the type of addressee, who could be a family member (nuclear or more distant), close friend or other acquaintance (cf. Bell 1984).

The mode and field dimensions of register are related to the concept of **genre**, here defined as a grouping of texts according to some conventionally recognised extralinguistic criteria (Lee 2001: 46). In this dissertation, I study variation in productivity within several genres, including personal correspondence, trial proceedings and the super-genres of casual conversation, imaginative texts and informative texts. As the majority of my studies are concerned with correspondence, a brief introduction to the letter genre is in order.

Nevala and Palander-Collin (2005: 2) define the **letter** as "a written message from one person to another" that has a function and several possible meanings depending on the recipient. The letter is also part of an intertextual network, being often a response to previous letters and giving rise to further responses. As noted above, an individual letter can deal with a number of topics, and the letter genre is heterogeneous in other ways as well. It has multiple subgenres, such as business letters, and letters can be both conventional and unconventional, formal and informal (Nurmi and Palander-Collin 2008: 25). Crucially to my research, Nevala and Palander-Collin (2005: 3–4) point out that "[t]he style in which people write letters may be correlated with their position in a group or society", and the style may also vary based on the recipient and the purpose of writing.

As noted by Nevala (2007: 89), letters were an integral means of communication between people in the 17th and 18th centuries. In this period, not all letters were necessarily seen as private: many letters had public functions and could be circulated or read aloud among family or friends (Nevala 2007: 89; Nevala and Palander-Collin 2005: 3). Thus, the writer had to keep in mind not only the recipient but possible overhearers. The early correspondence and other material used in this dissertation are described further in the next chapter.

4. Material

4.1. *Corpora of Early English Correspondence*

The *Corpora of Early English Correspondence* have been compiled at the University of Helsinki in a long-term project led by Terttu Nevalainen and Helena Raumolin-Brunberg. As the corpora have been designed for the purposes of historical sociolinguistics (see Chapter 3 above), the sampling unit has been the individual letter-writer, and the team have been aiming at a balance across various social categories, such as gender and social rank (Appendix II). Representing a "speech-like" genre (Culpeper and Kytö 2010: 17) and spanning c. 400 years from the 1400s to 1800, the corpora can be used to study sociolinguistics in the long diachrony. Amounting to roughly five million words in size, the corpora can be characterised as small but carefully compiled; for a discussion of the benefits of small corpora in diachronic research, see Hundt and Leech (2012).

Owing to a lack of resources in the compilation process, the *Corpora of Early English Correspondence* are based on published editions of letters rather than original manuscripts. However, the team have taken great care to only select reliable editions, and some of the editions have been checked against the originals and corrected by members of the team. For reasons of literacy and prestige, letters by women and the lower ranks have been harder to come by than letters by high-ranking men; therefore, the latter group is overrepresented in the corpora. Some poorer-quality editions with letters by women and the lower ranks have been used to supplement the corpora, and the quality of the edition has been coded into the header of each letter. A list of all editions used in the compilation of the corpora can be found in Nurmi et al. (2009: Appendix).

The present work makes use of all three of the correspondence corpora, described in Table 4.1 below: the original *Corpus of Early English Correspondence* (CEEC), c.1410–1681; the *Corpus of Early English Correspondence Supplement* (CEECSU), 1402–1663; and the *Corpus of Early English Correspondence Extension* (CEECE), 1653–1800. In addition, two alternate versions of the corpora are employed for the analysis of part-of-speech frequencies and key words, respectively: the *Parsed Corpus of Early English Correspondence* (PCEEC), which is the published, part-of-speech tagged and syntactically parsed version of the original CEEC, with some collections left out for copyright reasons; and the *Standardised-spelling Corpora of Early English Correspondence* (SCEEC), which currently comprises all of the collections in the correspondence corpora from 1500 onwards, with the spelling normalised using the VARD 2 software (Palander-Collin and Hakala 2011; Baron 2011).

More specifically, this dissertation uses the PCEEC in its entirety (Chapter 8); the 17th-century part of the original CEEC, supplemented with women's letters from the Betts and Thynne collections in the CEECSU (Chapter 6); the 17th-century collections of the original CEEC in the SCEEC format (Chapter 7); and the 18th-century parts of the CEEC and CEECE, from the year 1680 onwards (Chapters 10 and 11). These data sets are summarised in Table 4.2. At roughly one to two million words each, they are rather small; furthermore, in the analyses they are divided into sociolinguistically defined subcorpora whose size may vary between thousands and hundreds of thousands of words.

Table 4.1. The *Corpora of Early English Correspondence*

	CEEC	**CEECSU**	**CEECE**
Words	2,597,795	442,484	2,219,422
Letters	5,961	829	4,923
Writers	778	94	308
Time span	c.1410–1681	1402–1663	1653–1800

Table 4.2. The *Corpora of Early English Correspondence* used in this dissertation[1]

	PCEEC	**CEEC+ CEECSU C17**	**SCEEC C17**	**CEEC+ CEECE C18**
Words	2,157,573	1,323,945	1,223,846	2,216,119
Letters	4,969	3,172	3,055	4,945
Writers	659	324	305	315
Time span	c.1410–1681	1600–1681	1600–1681	1680–1800

4.2. *Old Bailey Corpus*

Another kind of material suitable for historical sociolinguistics is provided by the *Old Bailey Corpus* (OBC), which contains published trial proceedings from London's central criminal court. This is a "speech-based" genre (Culpeper and

[1] Note that the different versions of the CEEC corpora have been periodised slightly differently. Thus, the 17th-century collections of the SCEEC do not in fact contain all of the letters written in the 17th century in the corpus, as some of the 16th-century collections extend into the 17th century. Both this and slight differences in tokenisation (see further Chapter 13) contribute to the discrepancies between the sizes of the 17th-century materials reported in this table, as does the fact that the portion of the SCEEC used here does not include the supplementary material from the Betts and Thynne collections.

Kytö 2010: 17) enabling access to the language of the lower classes. However, the access is mediated through scribes and the changing editorial practices of the printer, which may affect the reliability of the results. For example, Huber (2007) finds a sharp decline in the frequency of negative contraction in the OBC over time, attributing this to the fact that in the 18th century, the proceedings became an official document controlled more and more by the City of London, which seems to have increased the formality of the reporting. Furthermore, scribes showed "differential faithfulness" (Huber 2007: 3.3.2.4) in their treatment of different linguistic variants, the faithfulness of the representation varying both intra- and inter-scribally.

Chapter 10 below uses version 0.4 of the OBC, which covers the years 1730–1910. The 18th-century part of the corpus, which is contrasted with the correspondence corpora in Chapter 10, contains c. 4.1 million words of speech-based material. However, unlike the correspondence corpora, which contain sociolinguistic metadata for almost all of the writers, the gender of the speaker in the 18th-century part of the OBC is known for c. 3.2 million words, and social rank in addition to gender for only half a million words. As court officials were men who used more formal language by virtue of their position, another subcorpus was constructed in Chapter 10 consisting of laymen and -women only. The subcorpora of the OBC used in Chapter 10 are described in Table 4.3.

Table 4.3. The subcorpora of the *Old Bailey Corpus* used in this dissertation

	OBC C18	**Gender known**	**Gender and rank known**	**Laymen and -women**
Words spoken	4,053,134	3,157,009	460,248	1,136,195
Speakers	25,340	22,471	2,555	8,072
Time span	1730–1800	1730–1800	1731–1794	1730–1794

In 2013, the developers of the OBC (Magnus Huber and team at the University of Giessen) published version 1.0 of the corpus, which provides considerably more data – and, presumably, metadata – than previous versions: c. 14 million words of speech-based material covering the period 1720–1913.

4.3. *British National Corpus*

The largest of the corpora used in this dissertation (Chapter 9), the *British National Corpus* (BNC) is a 100-million-word corpus of Present-day English compiled in the early 1990s. One subcorpus of the BNC in particular, the 4.2-

million-word demographically sampled spoken component, is well suited to sociolinguistic studies. Owing to the way the data was collected, however (see Burnard 2007: 1.5.1), metadata on the informants is patchy, and the subset for which both gender and social class are known is restricted to c. 2.6 million words (called BNC-DS in Chapter 9), which is comparable in size to the CEEC corpora and to the OBC. In terms of the kind of language included, everyday conversation, BNC-DS is more comparable to the correspondence corpora than to the trial proceedings, although correspondence may also include more formal language, such as business letters.

While the written component of the BNC was not designed with sociolinguistics in mind, it does contain some sociolinguistic metadata. Most importantly for the present study, the gender of the author has been recorded for c. 45 million words, which can be divided into imaginative and informative texts (Burnard 2007: 1.4.2). These two subcorpora (called BNC-W_{imag} and BNC-W_{inf} in Chapter 9) are large in comparison with the historical corpora, and not directly comparable to them in terms of genre. Most of the texts were written between the years 1985 and 1993, although some earlier works that were still in print in the early 1990s have also been included (Burnard 2007: 1.3–1.4). Due to the relatively short time span covered, the corpus is here treated as a synchronic snapshot of British English in the late 20th century. The spoken and written subcorpora of the BNC used in Chapter 9 are described in Table 4.4.

Table 4.4. The subcorpora of the *British National Corpus* used in this dissertation. The time spans are approximate as some of the texts are not given an exact date in the metadata

	BNC-DS	**BNC-W_{imag}**	**BNC-W_{inf}**
Words	2,632,512	15,931,189	29,322,653
Speakers/writers	358	445	889
Time span	1991–1993	1969–1993	1975–1994

5. Methods for diachronic corpus linguistics

To answer the research questions posed in Section 1.3 above, this dissertation employs a number of methods, which are described in detail in the individual studies. This chapter focuses on methods in two areas of general interest to diachronic corpus linguistics: statistical significance (Section 5.1) and visualisation (Section 5.2). Both sections begin with a survey of the state of the art, followed by an overview of the methods used in the present work. These methods will be evaluated further in Chapter 13 below.

5.1. Statistical significance

According to Coolidge (2013: 166), statistically significant findings "indicate that the results of the experiment are substantial and not due to chance". For significance to be measured, the experiment needs to be framed in terms of **hypothesis testing**, also known as significance testing. As an example, we may posit the research hypothesis that the frequency of a word in letters written in 1600–1639 differs from the frequency of the word in letters written in 1640–1681 – i.e., the frequency changes over time. We observe such a difference in a corpus and wish to know if the difference is significant or spurious. The **null hypothesis** would be that there is no relationship between time and frequency, so that an equally large difference could easily occur by chance in the corpus. The observed difference is statistically significant if the probability p that we are wrong in rejecting the null hypothesis is lower than a specific percentage, called the **significance level** (often 5%, $p < 0.05$). The procedure by which the probability is calculated is called a **significance test**.

5.1.1. State of the art

Historical corpora have traditionally been small (at most a few million words in size), which has sometimes made significance testing irrelevant or inapplicable, when there have only been a few instances of a linguistic feature per time period. Furthermore, the staples of statistical analysis in corpus linguistics, i.e., key words (Scott 1997; Rayson 2008) and collocations (Sinclair 1991; Evert 2009), are problematic to apply to historical corpora, which typically exhibit a great deal of spelling variation and which may span centuries of changing language use. A more common application of significance testing in diachronic corpus linguistics has been to analyse the significance of differences in the frequencies or pro-

portions of a limited number of specific linguistic features over time, combining the orthographic forms manually (e.g., Laitinen 2006).

With the advent of larger corpora, such as the 34-million-word *Corpus of Late Modern English Texts* 3.0 (based on Diller et al. 2010) and the billions of words contained in *Early English Books Online* (EEBO; Pumfrey et al. 2012), as well as new tools for spelling normalisation such as VARD (Baron 2011; Baron and Rayson 2009), the role of significance testing and other statistical methods is increasing in diachronic corpus linguistics. The present work shows that improvements in significance testing methods can also benefit smaller corpora.

The kind of significance tests that have been used the most in diachronic corpus linguistics, as well as corpus linguistics in general, can be characterised as **bag-of-words tests** (Chapter 7 below; Lijffijt 2013: Chapter 4). This means that when applied to comparing word frequencies, they make the assumption that words occur randomly in texts, which is obviously untrue and causes the tests to yield many **false positives**, i.e., spurious results marked as significant (Kilgarriff 2005; Paquot and Bestgen 2009; Lijffijt et al. forthcoming). As an upside, the tests are easy to apply: in comparing word frequencies between two corpora, for example, one only needs a 2×2 contingency table to represent the data, and the significance can be looked up in a book or by entering the numbers into an online calculator.

Examples of bag-of-words tests include the χ^2 test (Pearson 1900), the log-likelihood ratio test (Dunning 1993) and Fisher's exact test (Fisher 1922). While the χ^2 test is perhaps the easiest to understand and is still often taught first in courses on statistics, many corpus linguists have abandoned it in favour of the log-likelihood ratio test (Rayson and Garside 2000; Rayson et al. 2004), which has become the *de facto* standard test in key word analysis, widely implemented in corpus-linguistic software like WordSmith Tools (Scott 2012). Fisher's exact test, which is basically an exact version of the χ^2 test, used to be too computationally intensive for regular users, but nowadays it too is only a mouse click away, and has been recommended by, e.g., Pedersen (1996) for collocations. Nevertheless, the χ^2 test continues to be employed by many linguists for diachronic comparisons not involving key words or collocations.

Corpus linguists have long acknowledged the problem of the randomness assumption in these significance tests, and Gries (2008) proposes a solution to be used alongside the tests: a measure of dispersion, improved upon by Lijffijt and Gries (2012). Thus, even though the significance test itself does not take into account the dispersion of the item in question across the texts in the corpus, this is

done by a separate measure of dispersion, which can be used to evaluate the differences marked as significant by the test.

Another solution is to use a test that does take dispersion into account, henceforth called a **dispersion-aware test**. The input for these kinds of tests is somewhat more complex than a 2×2 table, as they typically require the relative frequency of the item in question to be reported for each text in the corpus, rather than for the corpus as a whole. These per-text frequencies are what enables the test to consider the dispersion of the item. Dispersion-aware tests include the t-test (e.g., Welch 1947); the Wilcoxon rank-sum test, also known as the Mann-Whitney U test (Wilcoxon 1945; Mann and Whitney 1947); and tests based on resampling, which are discussed in the next section.

Comparing the log-likelihood ratio test, the t-test and the Wilcoxon rank-sum test, Paquot and Bestgen (2009) find the t-test to be the best suited for their purposes. Kilgarriff (2001) recommends the Wilcoxon rank-sum test for comparing corpora, and it has recently been implemented in the online interface for BNC64, a socially balanced corpus of informal British speech extracted from the BNC (Březina 2013). While Vartiainen et al. (2013) use the Wilcoxon rank-sum test to compare pronoun frequencies over time and across social categories, dispersion-aware tests are still rare in diachronic corpus linguistics as well as the corpus-linguistic community as a whole.

The problems with bag-of-words tests have prompted some corpus linguists to propose the use of a measure of **effect size** in addition to a significance test (Gries 2005; Gabrielatos and Marchi 2012). If a difference is significant but its effect size is small, it means that the phenomenon is probably not spurious but that it is weak. Ideally, we would like to discover differences that are both real and large, so a measure of effect size might be a useful accompaniment to dispersion-aware tests as well.

Some researchers in diachronic corpus linguistics (see Hilpert and Gries forthcoming for an overview) are applying more sophisticated quantitative methods such as multifactorial regression analysis, which also involve statistical significance testing. These approaches are not dealt with in this dissertation, which focuses on so-called "robust statistics" (Hilpert and Gries forthcoming) requiring a minimal number of background assumptions. The two robust significance tests employed in the present work, permutation testing and bootstrapping, are discussed in the next section.

5.1.2. Resampling

The idea behind resampling statistics is to make the best use of the available data. For corpus linguistics, this means that the corpus is divided into samples, which are combined repeatedly and randomly to create confidence intervals for the observed frequency of an item in a subcorpus. This is a data-driven approach which in principle requires no modelling assumptions. However, the size of the sample matters: if we use individual words as samples, we make the invalid assumption that words occur randomly in texts (see Chapter 6 below), so it is usually better to use individual texts as samples, or, e.g., groups of texts written by the same person. Previous corpus-linguistic research has employed these types of statistics to verify the results of statistical modelling or to explore variability without hypothesis testing (Tweedie and Baayen 1998; Gries 2006b; Hinneburg et al. 2007), but this dissertation argues that they can also stand alone as useful measures of statistical significance.

5.1.2.1. Permutation testing

The impetus for the permutation testing approach in my work (Chapters 6, 9, 10 and 11) arose from a desire to compare the morphological productivity of a suffix across sociolinguistically defined subcorpora (based on, e.g., gender or time period) of different sizes. The measures of morphological productivity used in this dissertation, type and hapax frequencies, do not grow linearly with the size of the corpus, which means that they cannot be normalised, making comparisons difficult (see Chapter 6). With permutation testing, we can create confidence intervals for the type and hapax frequencies of the suffix in each subcorpus by comparing the subcorpus with multiple randomly composed subcorpora of the same size. This enables us to state how likely it is for a subcorpus to have the observed type or hapax frequency: for instance, if more than 99.9% of the randomly composed subcorpora have a higher type or hapax frequency than the actual subcorpus, the observed frequency is significantly low at $p < 0.001$.

The randomly composed subcorpora are built from the samples into which the entire corpus has been divided (see 5.1.2 above). These samples are picked in a random order by a computer program to compose a large number of subcorpora for each possible corpus size, from one word or token up to the size of the entire corpus. Each sample is used no more than once per subcorpus. Because there are usually too many possible combinations of subcorpora to compute exhaustively, we use a variant of permutation testing called Monte Carlo testing (Dwass 1957; Mitzenmacher and Upfal 2005: 252) to approximate the number of permutations

(see Chapter 6 for details). The permutation testing method is illustrated using type accumulation curves in Section 5.2.2 below.

5.1.2.2. Bootstrapping

The usual application of bootstrapping in diachronic corpus linguistics (Ogura and Wang 1996; Mannila et al. 2013) is as follows. We have a frequency measurement from a corpus and we wish to estimate the uncertainty in this frequency: If the corpus were composed slightly differently (but with the same number of samples), how much would the frequency of the item change? What are the limits to this variability? Bootstrapping (Efron and Tibshirani 1993) utilises a similar kind of resampling technique as permutation testing, except that the samples can be used more than once – it is random sampling with replacement. Thus, we can create random permutations of the entire corpus, rather than just a subcorpus, and get different frequency results every time, because we use some samples more than once, while others are left out entirely. These permutations are used to establish confidence intervals for the observed frequency in the same way as in permutation testing (5.1.2.1 above).

In my work with Jefrey Lijffijt and others in the DAMMOC project (Chapter 7 below; Lijffijt et al. forthcoming), we use bootstrapping as a significance test for comparing word frequencies between corpora, as in key word analysis. Say we have two corpora, *S* and *T*, and we are interested in the significance of the difference in the frequency of a word in them. For both corpora, we take the normalised frequency of the word in each sample (in this case, each text) and calculate the mean of these to derive a frequency estimate for the word in the corpus. Next, we estimate the variability in the mean frequencies by creating many randomised corpora having the same number of texts as the smaller of the two actual corpora. We do this for *S* and *T* separately: we repeatedly sample a corpus *S′* from *S* and a corpus *T′* from *T*, and calculate the mean frequency of the word in both *S′* and *T′*. Then we compute the *p*-value using a formula presented in Chapter 7 below. To put it in a slightly simplified way, we basically calculate the proportion of times that the frequency is greater in *T′* than in *S′* and multiply it by two, yielding the probability that the frequency differs in either direction between *S′* and *T′*. With some smoothing, this gives us the statistical significance we were after. Implementations of the test for R and Matlab are available in Lijffijt (2012).

As a concrete example of the use of this dispersion-aware test, let us take the word *Stephen* in the 17th-century collections of the standardised-spelling version of the CEEC, divided into two subperiods: 1600–1639 (consisting of 1,498 letters) and 1640–1681 (1,557 letters). *Stephen* occurs 25 times in 1600–

1639 but only 4 times in 1640–1681, and we hypothesise that this difference is significant at a level of $p < 0.05$. Let us test this hypothesis using the bootstrap test described above. The mean of the normalised frequencies of *Stephen* in each letter is 8.6 per million words in 1600–1639 and 3.9 per million words in 1640–1681. Now we use a computer program (Lijffijt 2012) to create ten thousand randomised corpora by repeatedly sampling 1,498 letters with replacement from both time periods, and calculate the frequency estimate for *Stephen* in each of the corpora. The proportion of times that the frequency is greater in the randomised corpora sampled from 1640–1681 than in those sampled from 1600–1639 is c. 30%. The program computes the *p*-value using the formula presented in Chapter 7 below, yielding the result $p \approx 0.6148$. This is well above our significance level, which means that the difference is not significant.

However, if we test the hypothesis using one of the bag-of-words tests, namely the log-likelihood ratio test, the difference is marked as highly significant at $p \approx 0.000007$. The explanation for this discrepancy is that the log-likelihood ratio test does not take into account the fact that the 25 instances of *Stephen* in 1600–1639 are poorly dispersed: in fact, all but one of them occur in a single letter written by William Wentworth in 1614. Crucially, the discrepancy between the tests is not limited to obscure names but also affects frequent words. In the analysis we conducted for Chapter 7, the difference in the frequency of the word *the* between the two periods was highly significant according to the log-likelihood ratio test ($p \approx 0.00000005$) but insignificant according to the bootstrap test ($p \approx 0.0173$) at a significance level of $p < 0.0016$. We used such a tight significance level because we tested a number of hypotheses at once, a situation which is discussed further in the next section.

5.1.3. Multiple hypothesis testing

All of the significance tests described above are meant for testing one hypothesis at a time. That is, the significance indicates the probability that we are wrong in rejecting a single null hypothesis. In diachronic corpus linguistics, however, we are often interested in testing multiple hypotheses within or across corpora – for instance, comparing item frequencies between each adjacent time period in the corpus. When we test multiple hypotheses, the significance should be adjusted for the number of hypotheses we are testing. To take a simple example, when flipping a coin once, the probability of getting heads is 0.5, but when we flip the coin four times, the probability of only getting heads is $0.5^4 = 0.0625$, and the probability of getting heads at least once is $(1 - 0.5)^4 = 0.9375$. Just like the probability of getting heads at least once increases when we increase the number of times we

flip the coin, the probability that we are wrong in rejecting the null hypothesis increases when we increase the number of hypotheses we test, unless we adjust the significance level accordingly.

There are several ways to correct for testing multiple hypotheses, none of which have been used much in corpus linguistics (Gries 2005: 281). In most of my own work on productivity, I have simply tried to keep the significance level sufficiently strict to ensure that the results are meaningful (Chapters 6, 9 and 10 below). Chapter 8 on comparing part-of-speech frequencies uses a significance test called Tukey's test (e.g., Coolidge 2013: 292–296), which is similar to the t-test but incorporates a built-in correction for **family-wise error rate** (FWER). The FWER metric is also implemented in a stand-alone method called Bonferroni correction, used in a corpus-linguistic setting by Oakes and Farrow (2007). According to Shaffer (1995: 569), adjusting significance through Bonferroni correction can be as simple as dividing the predetermined significance level by the number of hypotheses tested: for instance, if the significance level is 0.05 and we test 100 hypotheses, the corrected significance level is 0.05/100 = 0.0005.

Both Chapter 7 on comparing word frequencies and my final productivity paper in Chapter 11 use a method called **false discovery rate control** (FDR control), introduced by Benjamini and Hochberg (1995). We have also used it in Vartiainen et al. (2013). As noted by Lijffijt (2013: 14–15), while FWER expresses the probability that at least one false positive occurs, FDR is the expected proportion of false positives out of all positives. As discussed in Chapter 7 below, FWER is a more conservative metric and may thus mark interesting results as non-significant, whereas FDR is more powerful. FDR control is somewhat more complex than Bonferroni correction, but still easy to understand, and can be applied using a simple Excel spreadsheet or even pen and paper.

The procedure of FDR control is as follows (see Benjamini and Hochberg 1995: 293). Say we test ten hypotheses, a–j, each hypothesis stating that there is a difference in the frequency of a linguistic item between two adjacent time periods (so we have eleven time periods in all). We obtain *p*-values for the hypotheses using a significance test. Now we sort the hypotheses by *p*-value, smallest first, and number them from 1 to 10, as in Table 5.1. We choose a false discovery rate we deem acceptable; for instance, it could be 0.05 like the usual significance level. For each hypothesis, we divide the number of the hypothesis by the total number of hypotheses and multiply it by the false discovery rate. Then we compare the *p*-value to this new figure, let us call it the *q*-value, and look for the first row where the *p*-value is greater than the *q*-value. In Table 5.1, this occurs at hypothesis 6, so we shall only accept hypotheses 1–5. Without FDR control, we

would have accepted hypotheses 1–8 (at a significance level of 0.05). With the more conservative Bonferroni correction, the corrected significance level would be 0.05/10 = 0.005, so we would only accept hypotheses 1 and 2.

Table 5.1. An example of the FDR controlling procedure

	***p*-value**	**Number**	***q*-value**
Hypothesis f	0.001	1	1/10×0.05=0.005
Hypothesis j	0.004	2	2/10×0.05=0.010
Hypothesis d	0.010	3	3/10×0.05=0.015
Hypothesis h	0.012	4	4/10×0.05=0.020
Hypothesis g	0.024	5	5/10×0.05=0.025
Hypothesis c	**0.031**	6	6/10×0.05=**0.030**
Hypothesis a	0.033	7	7/10×0.05=0.035
Hypothesis b	0.046	8	8/10×0.05=0.040
Hypothesis e	0.060	9	9/10×0.05=0.045
Hypothesis i	0.070	10	10/10×0.05=0.050

5.2. Visualisation

According to Spence (2007: 5), the purpose of **information visualisation** is to allow information to be derived from raw data. The process is in principle straightforward: "data […] is transformed into pictures, and the pictures are interpreted by a human being", who gains insight or understanding in the process (ibid.). Visualisation facilitates understanding because human beings have evolved to "acquire more information through vision than through all of the other senses combined" (Ware 2004: 2). This section reviews the state of the art in diachronic corpus visualisation and describes the two major kinds of visualisation used in this dissertation, type accumulation curves and beanplots.

5.2.1. State of the art

Siirtola et al. (2011: Section 3) study the visualisations included in papers published in the journal *Literary and Linguistic Computing*. In diachronic corpus linguistics, the situation seems to be similar to that in the journal: typical visualisations include **line charts** and **bar charts** (Playfair 2005 [1786], 1801) showing the frequencies or proportions of linguistic items on the *y*-axis, with subcorpora representing different time periods on the *x*-axis (e.g., Leech 2011; Nevalainen 2013). Data of this kind is also sometimes presented using tables alone (e.g., Rissanen 2012). Line charts illustrating language change are often

roughly S-shaped; these **S-curves** have been modelled mathematically by, e.g., Blythe and Croft (2012).

Scatterplots (for the origins of which see Friendly and Denis 2005) are used to some extent in diachronic corpus linguistics to display per-text frequencies over time (e.g., Warner 2005). Another way to give some indication of per-text frequencies is to use **boxplots** (Tukey 1977: Section 2C), where the frequency of the item in each chronological subcorpus is given as a box rather than as a point (e.g., Tyrkkö 2014). As noted by Siirtola et al. (2011: Section 3.5), "[a] boxplot for a set of numbers shows the median, the first and third quartiles, and the smallest and largest values of the set visually". The boxplot is now being rivalled by the beanplot, discussed in Section 5.2.3 below. Hilpert (2011) presents one of the most innovative visualisations in diachronic corpus linguistics to date: interactive **motion charts**, which are series of diachronically ordered scatterplots animated using the googleVis package for R (Gesmann and de Castillo 2011). In a similar vein, Rohrdantz et al. (2011, 2012) track semantic change by various kinds of visual analytics. Lyding et al. (2012) use *Structured Parallel Coordinates* (Culy et al. 2011) to visualise change in academic discourse.

Other interactive visualisation tools of use to diachronic corpus linguistics include *DocuScope* (Kaufer et al. 2006) and *Text Variation Explorer* (Siirtola et al. 2014). DocuScope is a rhetorical analysis tool that has been used to compare and contrast Early Modern English texts, notably Shakespeare (Hope and Witmore 2010). Text Variation Explorer can be employed to compare (sub)corpora according to three common text measures (type/token ratio, the proportion of hapax legomena and average word length), as well as to cluster text fragments in terms of a user-specified list of words through principal components analysis. It is an innovative tool designed for exploratory corpus analysis, providing continuous and immediate feedback with all the graphs linked to each other and to the text itself to facilitate the generation of insights.

Some visualisations are used in connection with a specific method. **Dendrograms**, for example, are often used to illustrate the clusters produced by hierarchical clustering. Gries and Hilpert (2008) employ this method for periodisation in historical linguistics, whereas Tyrkkö (2013) applies it to group early English medical texts based on their most frequent words. The next section describes the visualisations associated with the method for measuring morphological productivity used in this dissertation, namely type accumulation curves.

5.2.2. Type accumulation curves

Figure 2.1 on page 18 above illustrates a single type accumulation curve, also known as a growth curve. In the permutation testing method used in this dissertation (Section 5.1.2.1 above; Chapters 6, 9, 10 and 11 below), we pick samples randomly without replacement to generate a million type accumulation curves using a computer program (Suomela 2007, 2014). Taking *-ity* types in the 17th-century section of the CEEC as an example, the procedure for constructing the type accumulation curves is as follows (curves for hapax legomena and tokens are constructed in a similar way; see Chapter 6 for details). First we divide the corpus into samples large enough to preserve discourse structure, e.g., individual texts. Then we pick a sample randomly and calculate the number of *-ity* types in it, plotting the sample on a figure with the size of the sample on the *x*-axis and type frequency on the *y*-axis. Next we pick another sample, add it to the previous one, and calculate the combined number of types, plotting the result on the same figure. We repeat this until all samples have been picked, producing a random type accumulation curve as in Figure 5.1.

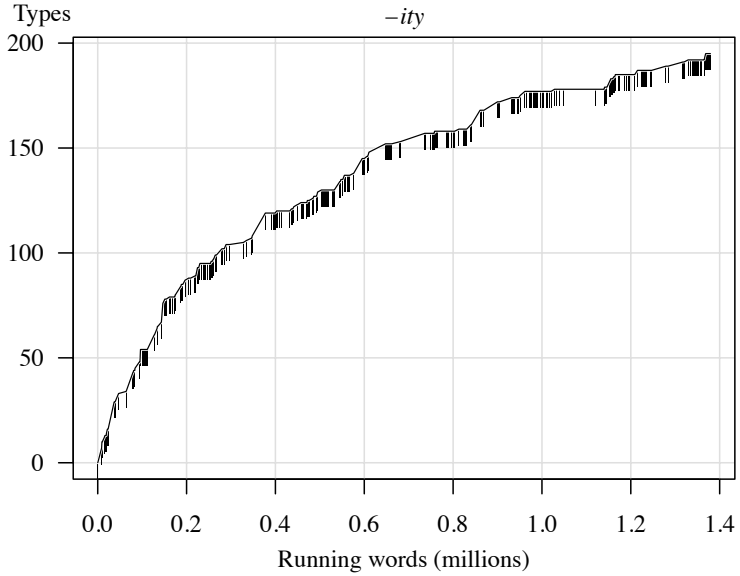

Figure 5.1. A random type accumulation curve (see Chapter 9 below). Each tick mark represents the addition of one sample

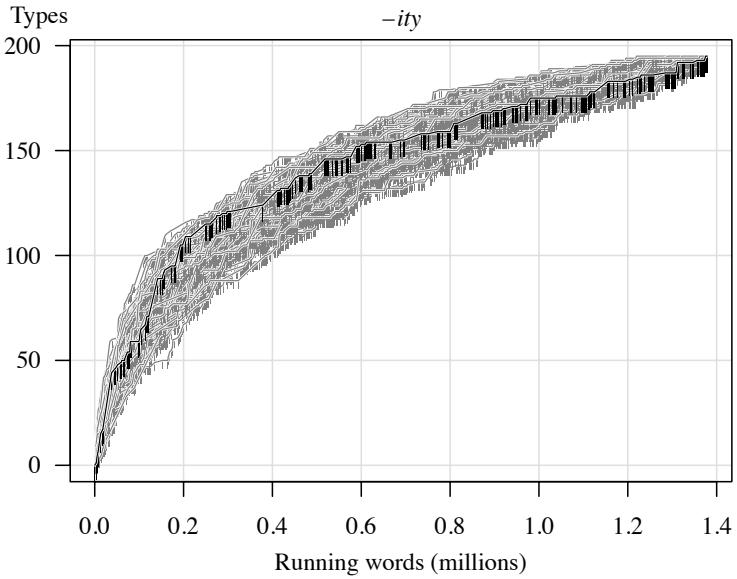

Figure 5.2. A hundred random type accumulation curves

We repeat this process multiple times, as in Figure 5.2, and keep repeating the process until we have plotted a million random type accumulation curves onto the graph. Then we draw confidence intervals onto the curves to indicate the area covered by 90%, 99%, 99.9% and 99.99% of the curves, shown in Figure 5.3. Now we can plot the desired subcorpora onto the curves and see where they are located on the confidence intervals. In Figure 5.4, the subcorpus of texts written in 1600–1639 contains a significantly low number of *-ity* types compared to randomly composed subcorpora of the same size ($p < 0.001$), whereas the subcorpus of texts written in 1640–1681 is not significantly different from random subcorpora of the same size ($p < 0.1$). This indicates that the productivity of *-ity* increases over time.

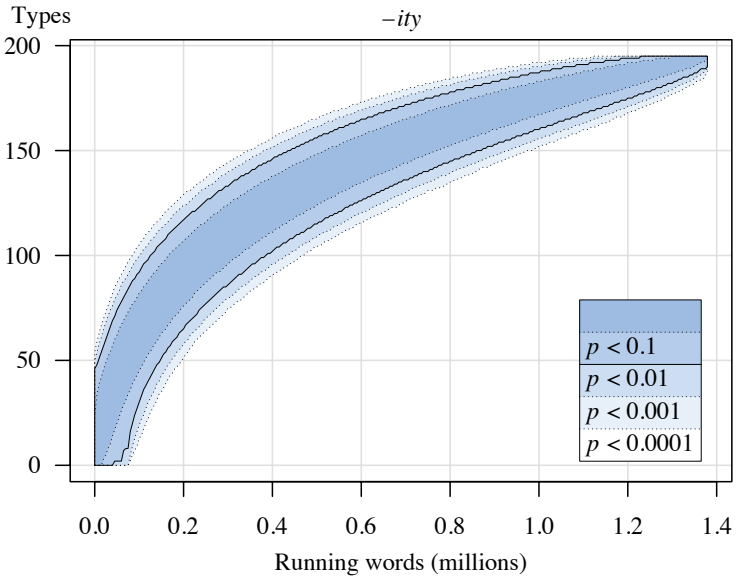

Figure 5.3. A million random type accumulation curves with confidence intervals (see Chapter 6 below)

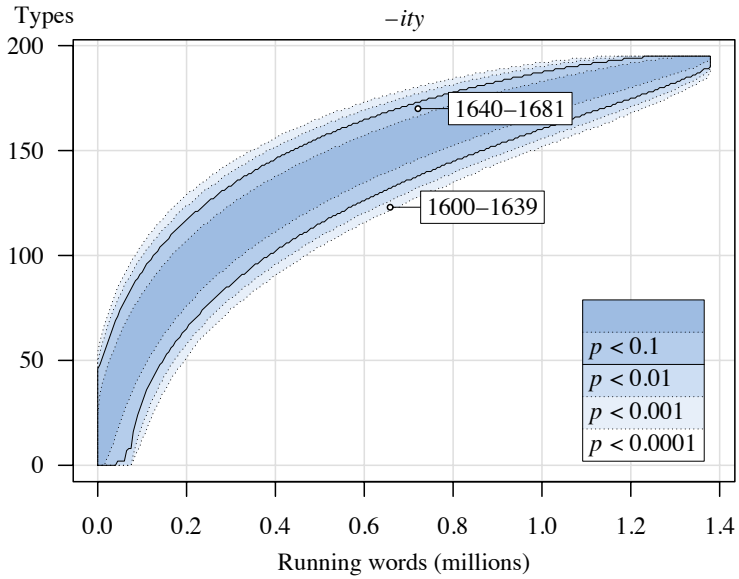

Figure 5.4. Subcorpora of time periods plotted onto a million random type accumulation curves with confidence intervals (see Chapter 6 below)

Note that unlike most graphs depicting change over time, the type accumulation curves do not have time on the *x*-axis: rather, the *x*-axis shows corpus size, measured in either the number of running words or the token frequency of the linguistic item in question. We can only use this graph to illustrate change over time if we plot chronological subcorpora onto it. The fact that the subcorpus for 1640–1681 is plotted to the right of the subcorpus for 1600–1639 does not indicate time but corpus size, as the 1640–1681 subcorpus happens to be the larger of the two. Furthermore, note that the *y*-axis shows the *absolute* number of types (recall that type frequency cannot be normalised), so the fact that the subcorpus for 1640–1681 is vertically higher on the graph than the subcorpus for 1600–1639 does not necessarily indicate that it has a *relatively* higher type frequency. That is, we cannot compare these two subcorpora with each other directly; we can only compare each subcorpus with randomly composed subcorpora of the same size. For more discussion on the use of type accumulation curves in diachronic corpus linguistics, see Chapter 10 and Chapter 13 below.

5.2.3. Beanplots

As noted in Section 5.2.1 above, the beanplot (Kampstra 2008) is an improved version of the boxplot. As far as I have been able to determine, Chapter 8 below (i.e., Säily et al. 2011) represents the first use of the beanplot in diachronic corpus linguistics. The chief difference between beanplots and boxplots is that the beanplot shows the actual shape of the frequency distribution of the item in question across the samples (Figure 5.5), whereas the boxplot is simply box-shaped (Figure 5.6). In some ways, however, the boxplot is more sophisticated, since it explicitly marks the first and third quartiles and the smallest and largest values, as well as giving an indication of statistical significance: if the wedges in the middle of two boxplots do not intersect, the frequency of the item differs significantly between the two. The idea behind the beanplot is to be more accessible to non-expert users, so advanced features like these have been left out.

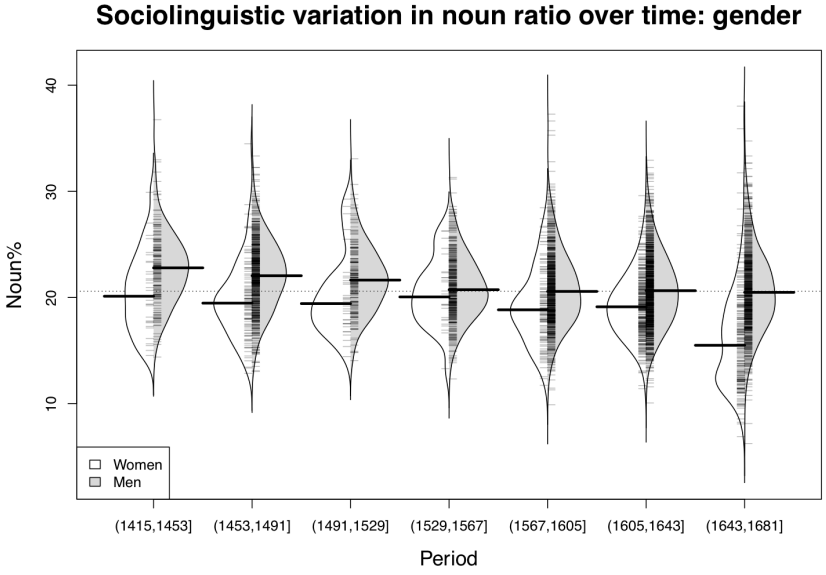

Figure 5.5. A beanplot illustrating gender-based variation in the proportion of nouns in the *Parsed Corpus of Early English Correspondence*. Reproduced from Siirtola et al. (2011: Figure 12)

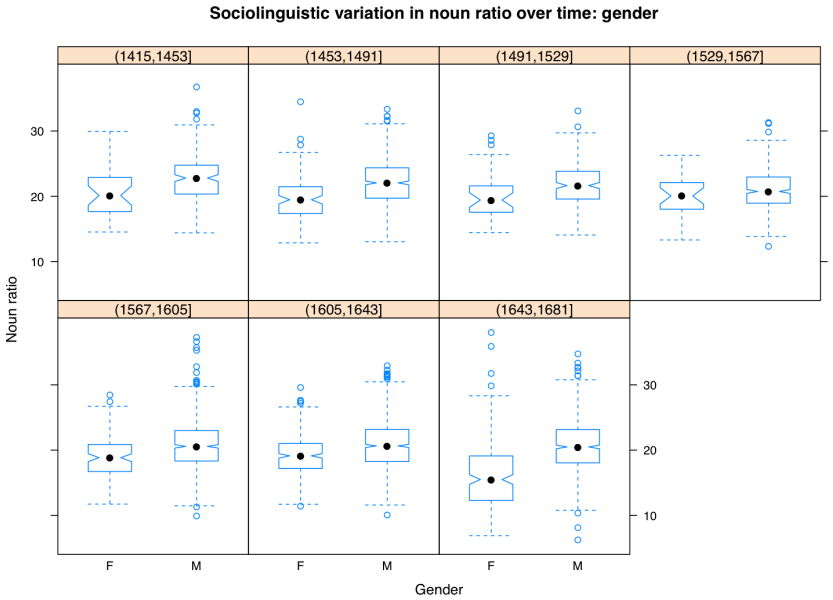

Figure 5.6. Boxplots illustrating gender-based variation in the proportion of nouns in the *Parsed Corpus of Early English Correspondence* (F=women, M=men). Reproduced from Siirtola et al. (2011: Figure 11)

The beanplot consists of a one-dimensional (vertical) scatterplot accompanied by a density plot (e.g., Silverman 1986) showing the shape of the distribution. The left-hand side of the beanplot can represent a different subcorpus from the right-hand side, as in Figure 5.5, or the density plot and scatterplot can be mirrored to form the bean. Each tick mark on the scatterplot represents the frequency of the item in one sample (here, one letter), so the number of tick marks gives an indication of the size of the subcorpus. The thick horizontal line on the scatterplot represents the mean frequency of the item across the samples in the subcorpus, while the mean frequency across all samples is represented by a dotted horizontal line across the entire graph. Note that this differs from the boxplot, which uses median frequency; however, this can be changed, and Vartiainen et al. (2013) indeed use beanplots with median frequencies.

The density plot facilitates spotting multimodality in the data, which can be caused by outliers (see Siirtola et al. 2011: Section 3.5). These are visible in the form of "bumps" on the bean; for instance, the bump on the lower left-hand side

of the last bean in Figure 5.5 is due to letters written by Dorothy Osborne, a gentlewoman who has turned out to be an outlier in terms of nearly every linguistic feature studied in the CEEC corpora. Here she uses fewer nouns than most of the other women in the PCEEC, causing the mean proportion of nouns in women's letters to drop considerably lower in her period than in the other periods. In the boxplot in Figure 5.6, the median is also low in the final subcorpus of women's letters, but the outlier dragging it down is not visible on the plot in any way. Thus, the beanplot is in my opinion preferable to the boxplot (as well as to the simple line and bar charts) for making visual comparisons in diachronic corpus linguistics, where corpora are typically small and individual outliers may have a large impact. An implementation of the beanplot for R is available in Kampstra (2008).

PART III: CONCLUSION

12. Answers to research questions

1. Is there sociolinguistic variation and change in the productivity of *-ness* and *-ity* in the history of English?

We have charted sociolinguistic variation and change in the productivity of *-ness* and *-ity* from Early Modern English to Present-day English. In short, the answer to the research question is yes, except that the productivity of *-ness* does not seem to change over time. As hypothesised, we have discovered that the productivity of the etymologically foreign *-ity* is significantly low with women in the 17th century (Chapter 6). Surprisingly, however, the difference seems to persist throughout the centuries and across genres, although in 18th-century correspondence it is only apparent in the register of letters written to close friends (Chapter 11). This has led to a new hypothesis of a relatively stable gendered style.

We have also hypothesised that the lower ranks would use *-ity* significantly unproductively in the 17th century, but there has been too little data from them for significant results to emerge. In 18th-century correspondence, we do find that the lower-class Clift family underuse both *-ity* and *-ness* compared to the royalty, which could be due to fact that the topics they write about do not require extensive use of abstract nouns (Chapter 11). Furthermore, in present-day conversations in the BNC (Chapter 9), the productivity of *-ness* is significantly low with women belonging to the lower classes (while the productivity of *-ity* is low with women in general), which could be a question of style. This shows that social categories may co-vary, which should be taken into account where the amount of data permits.

The above already disproves our hypothesis that there is no sociolinguistic variation in the productivity of *-ness*. Moreover, in 18th-century correspondence, there is a tendency for clergy to use *-ness* more productively than others, perhaps as a carry-over from sermons, whereas royalty use it significantly unproductively by repeating the same tokens in formulaic expressions (Chapters 10 and 11). Nevertheless, there seems to be less variation in the productivity of *-ness* than in that of *-ity*: even though the same categories have been tested for each, there are fewer statistically significant differences in the use of *-ness*, and the most significant differences are in the use of *-ity*. Hence, the productivity of *-ness* also seems to be more resistant to change in the periods studied.

Our exploratory analysis has shown that the productivity of *-ity* increases throughout the 17th and 18th centuries, perhaps led by the social rank of

professionals and, in 18th-century correspondence, by men writing to their close friends (Chapters 6, 10 and 11). This could be connected to the overall long-term increase in the productivity of *-ity* observed by Lindsay and Aronoff (2013) in the *Oxford English Dictionary*, but it could also be associated with stylistic changes in the two genres studied, personal correspondence (Biber and Finegan 1997) and courtroom discourse (Huber 2007).

2. How can we study productivity in small corpora which contain a great deal of spelling variation?

The method presented in Chapter 6 seems to be a good solution to this if we value data-driven exploratory analysis without complex background assumptions. Even though the method does require a certain amount of data to yield reasonable confidence intervals, we have been able to discover both significant and interesting differences in corpora containing only a million words or so, divided into various sociolinguistically defined subcorpora. Spelling variation is a minor issue because we do not attempt to track the frequency of the base of each formation or the frequency of all hapax legomena in the entire corpus; rather, we concentrate on the type and hapax frequencies of the affix as a function of either the number of running words or the number of affix tokens in the corpus. This gives us an improved version of the productivity measures *V* and *P* (Baayen 2009) without the problem of comparability across subcorpora of varying sizes, as we use the statistical technique of permutation testing to only compare each subcorpus with randomly composed subcorpora of the same size.

3. How can we study variation and change in corpora which may not be completely comparable over time and across genres?

Chapter 7 proposes two related methods for comparing corpora and studying genre continuity. The first of these is the bootstrap test (Section 5.1.2.2 above), which is a reliable and data-driven way to determine the significance of differences in **word frequencies**, combined with false discovery rate control (see 5.1.3 above). The second method is useful when we are interested in differences related to a particular semantic category. The idea is to go through the words whose frequency is significantly different between two corpora according to the bootstrap test and narrow the list down to only those words that belong to the category. Rather than defining the category in an ad-hoc manner as has been done in previous research (Leech and Fallon 1992), we use the *Historical Thesaurus of the Oxford English Dictionary* (HT) to classify the words. Another solution

would be to tag the corpora semantically and identify key semantic domains (Rayson 2008) using the bootstrap test, but no current semantic tagger is completely accurate or uses a classification scheme sensitive to change over time like the HT (cf. Archer 2012, 2014).

Chapter 8 uses beanplots (Kampstra 2008) to visually compare differences in **part-of-speech frequencies** across time periods and social categories. As discussed in 5.2.3 above, beanplots are preferable to boxplots and simple line graphs in that they show the shape of the frequency distribution across the texts in each subcorpus, which facilitates identifying outliers. Furthermore, beanplots give an indication of the number of texts in each subcorpus, enabling the researcher to assess the importance of the observed differences more accurately. The statistical significance of the differences is measured by Tukey's test, which provides a built-in correction for testing multiple hypotheses. Other dispersion-aware tests could be used as well (see 5.1 above). To assess the reliability of the part-of-speech tagging of the corpus, we reannotate the corpus according to a different annotation scheme and compare the part-of-speech frequencies in the two versions.

These studies illustrate the extent to which different sections of the *Corpus of Early English Correspondence* vary in terms of word frequencies as well as noun and pronoun frequencies. I would argue that large-scale studies of this kind can be extremely useful to researchers in diachronic corpus linguistics, most of whom study the frequencies of a small subset of words or constructions. It is up to the individual researcher using the corpus to decide whether the variation and change observed in the large-scale studies could affect her results. For instance, the relative stability of the corpus over time observed in both Chapter 7 and Chapter 8 gives me a solid basis on which to build my observations of change in morphological productivity. Furthermore, the finding from Chapter 8 that men consistently use more nouns than women is highly relevant to my research and affects my interpretation of the gender-based variation in the use of the nominal suffix *-ity* in the corpus.

4. Are the productivity measures proposed in previous research valid in and applicable to sociolinguistic data of this kind?

Chapters 6 and 9 discuss reasons why the productivity measures V and P cannot straightforwardly be applied to comparing sociolinguistically defined subcorpora of varying sizes. While some studies have solved this problem by using parametric models (Plag et al. 1999; Keune et al. 2006), these entail simplifying assumptions about the text, e.g., that the occurrences of the words are in-

dependent. The solution presented by Gaeta and Ricca (2006), to reduce the size of the larger subcorpora to match the smallest subcorpus, requires us to discard some of the already sparse data. Chapter 6 proposes a robust, assumption-free approach that provides access to the actual growth curve instead of an estimation, makes the most of the available data and incorporates a reliable measure of statistical significance.

Chapter 9 finds that the applicability of hapax-based measures in general depends on the size of the corpus. This is not due to the fact that hapaxes are less likely to be genuinely new formations in small corpora; rather, the reason is that the number of hapaxes varies so widely in small corpora that it is impossible to tell whether a certain number of hapaxes is typical or significantly different from the norm. The good news is that the method presented in Chapter 6 enables us to visualise the amount of variability in hapax frequencies in the corpus and to leave out such measures if necessary. Unfortunately, it seems that at around a million running words, many sociolinguistic and historical corpora are too small for hapax-based measures. Chapter 9 concludes that hapax-based measures remain theoretically valid and that in studies of variation and change in productivity, hapax accumulation curves should be used in addition to type accumulation curves when the corpus is large enough.

5. What are the requirements for a usable tool for studying variation in productivity in data of this kind?

As noted in Chapter 10, exploratory analysis of variation and change involves testing a number of hypotheses. This gives rise to two requirements: firstly, we need a way to adjust the significance level for multiple hypothesis testing, and secondly, we need a convenient way to browse through the results. A requirement specific to our method is the need to conveniently switch the measure of corpus size on the *x*-axis from the number of running words to the number of suffix tokens, as both of these measures have proved to be relevant and to sometimes yield different results.

The new version of our software (Suomela 2014) provides actual *p*-values in addition to confidence intervals and thus a way to correct for testing multiple hypotheses, which is done automatically by the program using a method called false discovery rate control (see Section 5.1.3 above). Furthermore, the curves are implemented as interactive SVG images embedded on web pages with links to the other images and to the underlying data. In addition, the results can be viewed in an SQLite database, which also holds the input data and can be queried in various ways.

While the initial setup is done via the dreaded command-line interface (cf. Garretson 2008: 80), after the program has finished running, everything can be viewed using the familiar web browser interface and/or the SQL database tool of the user's choice. The figures are also provided as high-quality static PDF images that can be embedded in publications. Thus, we take into account the observation by Theus and Urbanek (2008: 5–6) that different kinds of graphics are needed for exploration on the one hand and presentation on the other.

13. Evaluation of methods

This chapter presents a critical evaluation of the main methods used in the studies, divided into three categories: morphological productivity (Section 13.1), statistical significance (Section 13.2) and visualisation (Section 13.3). As a sort of prelude, I would like to take up the matter of preprocessing the material used. In my collaboration with computer scientists, I have come to notice that they all have their own preferences as to how the input data is preprocessed for analysis. In some cases, this has led to differences in word counts of the same corpus, as tokenisation is performed in different ways by different programs. In most cases, however, we have used the official word counts provided with the corpus, when those have been available. Tokenisation becomes an issue especially in comparisons of normalised frequencies across corpora, e.g., when we wish to compare part-of-speech frequencies across corpora in the long diachrony (see Chapter 8).

13.1. Morphological productivity

13.1.1. Comparing type counts

Most of the results regarding the morphological productivity of *-ness* and *-ity* in this dissertation were obtained using the measure of type frequency, which Baayen (2009) calls realised productivity and which is only one of his three facets of morphological productivity. The other two are the hapax-based *P* and *P** measures, which estimate the growth rate of the affix and the contribution of the affix to the growth rate of the total vocabulary, respectively. It could be argued that if hapax-based measures cannot be used in small corpora, we are effectively only measuring lexical richness rather than morphological productivity.

As noted in Chapter 10, however, in diachronic corpora the aspect of productivity concentrating on new formations is not completely neglected, as we have access to change over time in the realised productivity. Furthermore, we can try to limit the types we count to productive uses of the affix only. For instance, in my MA thesis (Säily 2008), which studied productivity in the 17th-century section of the CEEC, I experimented with leaving out *-ness* and *-ity* types that did not have an extant base, that were prefixed or that had been in the language for more than a century according to the *Oxford English Dictionary*. As the results were similar in each case, even with all three restrictions in place, I decided to be as inclusive as possible in my dissertation and to count all nouns that etymologically contained the suffix (see further Chapters 6 and 9). In Chapter 9, I

analysed the effect of the restrictions in the spoken demographic section of the BNC and again found that the results were similar in each case.

Another way to limit the kinds of types counted in a diachronic corpus to productive uses only is to study the number of new types, i.e., the chronologically first occurrences of the types in the corpus, possibly with a starting lexicon of some kind (e.g., Cowie 1999; Gardner 2013). This can also be accomplished using my method by doing pairwise comparisons, as in Chapter 6, which splits the corpus into two periods and plots them on the curves. The subcorpus representing the first period is situated low on the curves, which means that the remaining texts, which are all from the second period, add a large number of new types to those of the first period. A starting lexicon could be used to exclude old types before such comparisons.

Perhaps a more serious issue with sociolinguistic studies of productivity is that there is often not enough data to reliably analyse combinations of categories and thus no way to conclusively determine, e.g., which social group is leading the change. The method of permutation testing used in the present work errs on the side of caution, meaning that with a small amount of data, gaining statistically significant results may be impossible. Nevertheless, I prefer robustness and reliability to possibly spurious results. Near-significant tendencies can be discovered through a visual inspection of the type accumulation curves. Furthermore, quantitative studies can be complemented with qualitative research on the behaviour of individuals in their social contexts, giving us a richer picture of the variation (see Chapter 11).

Large corpora, too, have their problems. For instance, there may be so many hits that preprocessing the data demands a great deal of resources, as many of the types – especially the hapax legomena – are misspellings or variant spellings that need to be subsumed under the correct lemma. The issue becomes particularly pressing with large historical corpora, such as the promising new EEBO-based corpus currently under development at Lancaster University. As spelling standardisation can usually only be applied to sufficiently frequent items, most of the types may remain unstandardised, leaving the researcher with thousands of types (or millions of tokens) to lemmatise and categorise manually.

13.1.2. Hapax legomena in small corpora

As noted above, hapax-based measures of productivity proved to be unusable in small corpora. However, the amount of variability in the frequency of hapax legomena does not only depend on corpus size (measured in either the number of running words or the number of affix tokens) but also on the affix in question. It

seems that there is more variability in the number of *-ity* hapaxes than in that of *-ness* hapaxes in small corpora. Figure 13.1 shows that the bounds for hapax legomena are wider for *-ity* than for *-ness* in the 17th-century section of the CEEC, and the same phenomenon occurs in each of the corpora studied in this dissertation with the exception of the large written subcorpora of the BNC.

Furthermore, looking at either hapax or type accumulation curves for *-ness* and *-ity* in any of the corpora used in this dissertation, it seems that the curves for *-ity* rise more steeply than those for *-ness* at small corpus sizes, as in Figure 13.1 and Figure 13.2. This may mean that *-ity* has more high-frequency types that are often encountered soon after the beginning of the corpus (these are also initially counted as hapaxes until the size of the corpus grows and more instances of the same type are found). Figure 13.3 confirms this for the 18th-century section of the CEEC corpora. The most common *-ity* types in the data set are *opportunity*, *society*, *necessity*, *university*, *quantity*, *curiosity*, *quality*, *civility*, *authority* and *sincerity*. The first of these is typical of correspondence as people often mention having the *opportunity* to write; some are entities such as *society* and *university*, while others are more like embodied attributes, e.g., *curiosity* and *sincerity*, the latter of which may appear in closing formulae.

It is unlikely that either the wideness of the bounds or the initial steepness of the curves is due to my inclusive policy on what counts as an *-ity* word, as the same phenomena also occur when the restrictions mentioned in Section 13.1.1 are in place, in both the 17th-century section of the CEEC and the spoken demographic section of the BNC. What, then, could the reason be? On the one hand, it could have something to do with the fact that the use of *-ity* is marked both socially and stylistically, which causes the occurrences of *-ity* to be poorly dispersed in the corpus. In other words, people are extremely divided in how productively they use *-ity* depending on both their social background and the situation of use, and in small corpora these individual differences mean that the bounds for hapax legomena become wide. On the other hand, it could be that regardless of dispersion, the lower productivity of *-ity* as manifested in the low number of hapaxes at the level of the entire corpus and in the large number of high-frequency types is the reason for these phenomena.

Figure 13.1 and Figure 13.2 show that the shape of type and hapax accumulation curves is affix-specific and can reveal details of the behaviour of the affixes. Thus, hapax accumulation curves can be useful even in small corpora, not for discovering statistically significant differences in the productivity of an affix across subcorpora but as a heuristic tool for comparing different affixes.

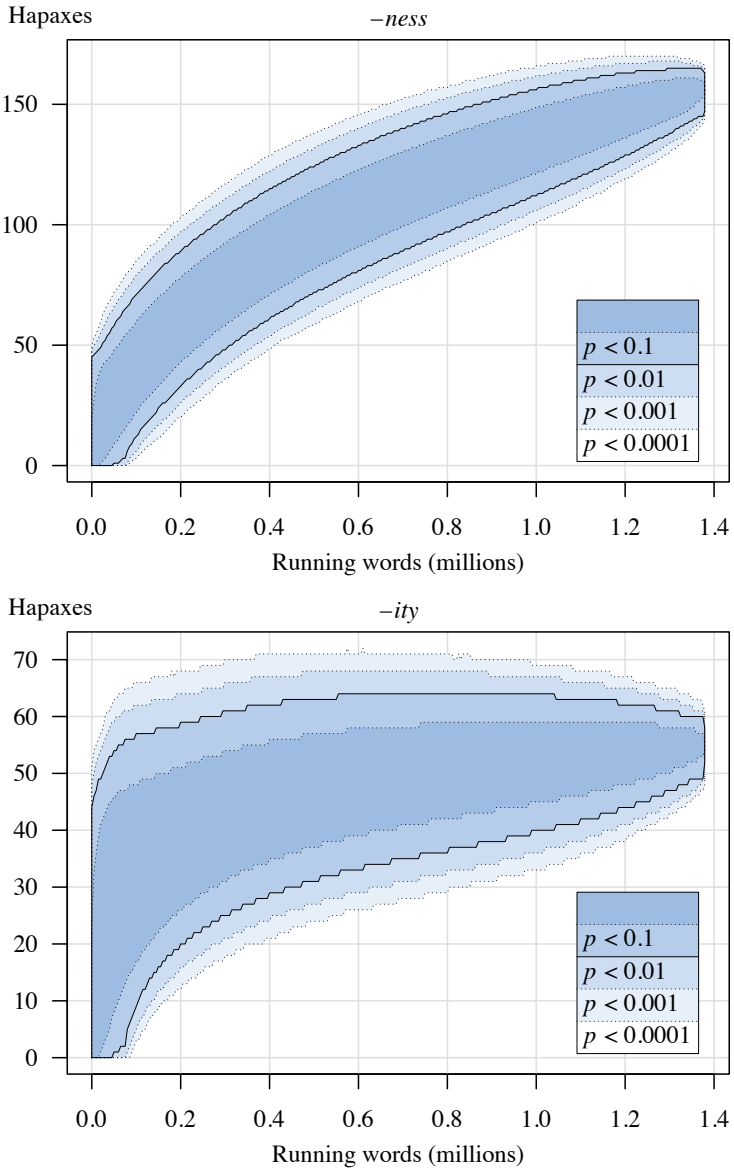

Figure 13.1. Hapax accumulation curves for *-ness* (top) and *-ity* (bottom) in the *Corpus of Early English Correspondence*, 1600–1681 (see Chapter 6 above)

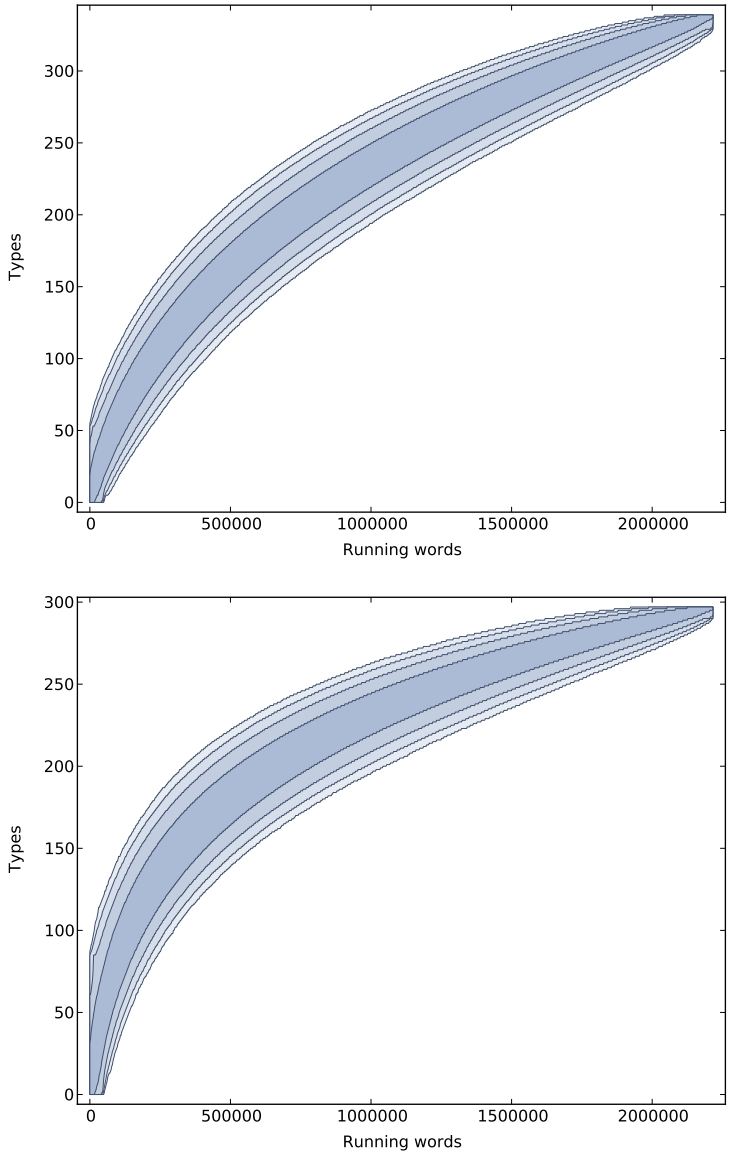

Figure 13.2. Type accumulation curves for *-ness* (top) and *-ity* (bottom) in the *Corpora of Early English Correspondence*, 1680–1800

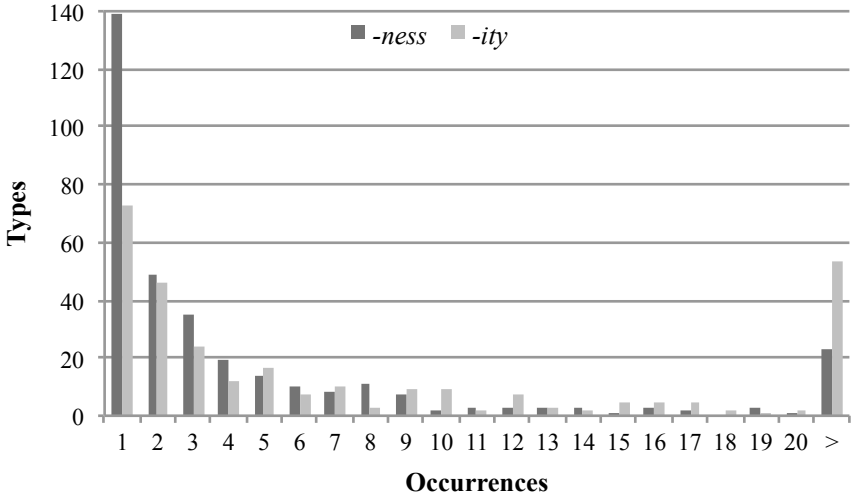

Figure 13.3. Frequency spectra of *-ness* and *-ity* types in the *Corpora of Early English Correspondence*, 1680–1800. There are more hapax legomena in *-ness* than in *-ity*, but there are more high-frequency types (occurring more than 20 times) in *-ity* than in *-ness*

13.2. Dispersion-aware tests

The choice of which test to use for estimating statistical significance is partly a matter of the researcher's preference. When comparing (normalised) word frequencies across corpora or subcorpora, any of the dispersion-aware tests mentioned in Section 5.1.1 above can be used. The tests based on resampling are basically assumption-free and make the most of the data, while the **t-test** entails the assumption that the mean frequencies follow normal distributions, which may not apply to all cases. The **Wilcoxon rank-sum test** does not look at the frequencies themselves but at the ranks of the samples ordered by frequency, which is a coarser measure and has been criticised by, e.g., Rayson and Garside (2000: 2) for discarding "most of the evidence we have about the distribution of words". Nevertheless, it has been used with success by, e.g., Vartiainen et al. (2013), and it outperforms the t-test in a comparison by Lijffijt et al. (forthcoming). As noted in Section 5.1.1, it has been implemented in Březina (2013), and it is arguably easier to carry out than tests based on resampling.

This dissertation concentrates on tests based on resampling because they are robust and make the best use of the available data, which is especially important considering the bad-data problem in historical linguistics (Labov 1994: 11). Even though **permutation testing** is here used to discover significant differences in morphological productivity (type and hapax frequencies) across subcorpora, it can be used to find significant differences in the type, hapax and token frequencies of any linguistic item. The advantages of this method include its lack of simplifying assumptions as well as its highly visual nature, discussed further in 5.2.2 above. It is also now readily available, implemented in Suomela (2007, 2014), and has already been used by Gardner (2013) and Bentz et al. (forthcoming). A possible disadvantage is that it may be overly conservative in some situations (Lijffijt 2013: 35, 38), making it difficult to obtain significant results when analysing small corpora or rare features.

As for choosing between permutation testing and **bootstrapping**, the choice depends on the kind of frequencies we wish to compare. When comparing token frequencies, either test can be used (see Lijffijt 2013: 29ff.). Like the permutation testing software (Suomela 2014), the implementations of the bootstrap test for R and Matlab (Lijffijt 2012) are relatively straightforward to run but do require some familiarity with the program. It could be said that bootstrapping assumes slightly more than permutation testing because the former employs random sampling with replacement: if we are prepared to use the same sample multiple times to estimate uncertainty in the frequency, we must assume that the corpus could in principle contain multiple samples that are very similar to each other, and thus assume something about the world beyond the corpus. These assumptions are commonly accepted in corpus linguistics, as we generally do think that the corpus is representative of a target population, and that there are groups of texts that are similar to one another for a variety of reasons.

When comparing type frequencies, however, only permutation testing is applicable. This is because bootstrapping will always underestimate the upper bounds for type frequencies. When the same sample is used more than once, we will not get any new types from it, because the same types will already have been observed the first time the sample was used. Thus, the estimated number of types will never exceed the original number of types.

Resampling approaches have been criticised for being computationally intensive. For example, Bestgen (2014) argues that permutation testing would be infeasible if we tested a very large number of hypotheses (e.g., key word analysis across several corpora) and simultaneously wished to have a very strict significance threshold. I would say that for most applications, resampling is quite feasi-

ble using today's computational equipment (see Suomela 2014: Performance and Scalability). Sometimes you might need to wait for a day for the results, but if they are much more reliable than those yielded by other tests, they will be worth the wait. A rough idea of the results can usually be gained within a few minutes by lowering the number of randomisations computed, and in the case of bootstrapping, Lijffijt (2013: 30–31) shows that the time-consuming randomisation process can sometimes be avoided altogether if the samples are large enough. At any rate, strict significance thresholds may have been used in the past mostly to cope with the multitude of false positives yielded by the bag-of-words tests, so there may be no need to be quite so strict with robust dispersion-aware tests, especially when applying post-hoc correction for testing multiple hypotheses, discussed in Section 5.1.3 above.

13.3. Visualising variation and change

13.3.1. Beanplots and outliers

Beanplots are an excellent way to visually compare normalised frequencies over time and across subcorpora. The fact that they facilitate identifying outliers through the shape of the frequency distribution across the samples is not only useful in small corpora but also in large and messy data sets, where we might not know what genres they represent and what other kinds of variation there might be. This applies to both historical materials like EEBO and, e.g., materials collected from the Internet.

Another way to combat outliers in combination with beanplots is to choose the sample wisely. As noted in Section 5.2.3, the frequency of the linguistic item in the (sub)corpus as a whole is calculated by taking the mean or median of the normalised frequencies in the samples. In Section 5.2.3 we used individual texts as samples, and it was for these per-text frequencies that the shape of the frequency distribution was drawn. Having identified an outlier as illustrated in that section, we can remove her data from the analysis, as in Figure 13.4. In a sociolinguistic setting, however, it might make more sense to avoid the problem by using a different kind of sample, namely, a group of texts written by an individual person. The group could represent, e.g., all texts written (or spoken) by that person, all texts written by the person during a specific period of time, or all texts written by the person to a specific (type of) recipient. In this way, each person would not contribute more than a few samples, so individual outliers – even those with a large number of texts in the corpus – would only have a minor impact on the results. This has been done by Vartiainen et al. (2013), and it is also

the approach I have taken in my studies of morphological productivity using type accumulation curves (Chapters 6, 9, 10 and 11).

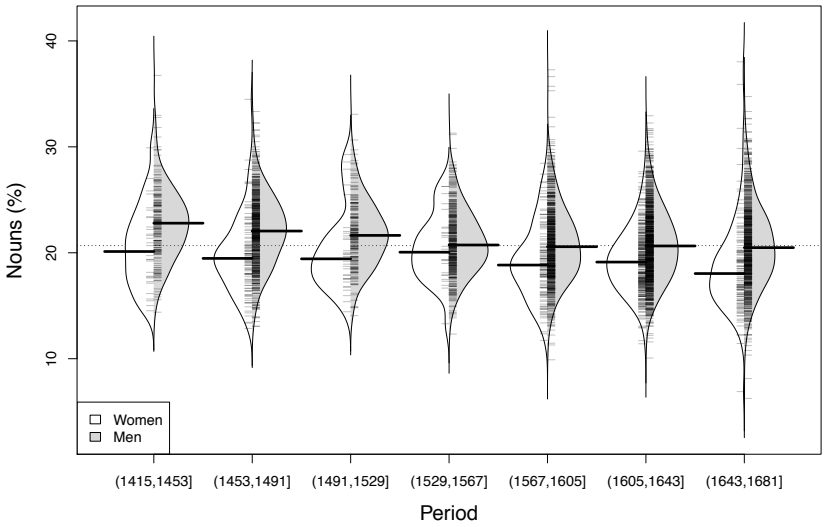

Figure 13.4. A beanplot illustrating gender-based variation in the proportion of nouns in the *Parsed Corpus of Early English Correspondence*. Letters by Dorothy Osborne have been removed from the final period (cf. Figure 5.5; see Chapter 8)

13.3.2. Towards interactive visualisation

Beanplots, as well as any other kind of visualisations used in diachronic corpus linguistics, could be improved through the addition of interactive features enhancing exploration (Pike et al. 2009; Siirtola et al. 2011: Section 5). Most importantly, the tick marks of the scatterplot in the middle of the beanplot could be linked to the underlying data for easy access to both the frequencies and the actual texts on which the frequencies are based. In addition, it would be helpful to be able to change the periodisation on the fly by moving a slider or by typing in the desired length of the periods. The selection of the subcorpora to be compared, too, would be more convenient through a graphical interface. However, I am not sure whether these features could be added to the current implementation of the beanplot in R (Kampstra 2008). While there are some programs providing inter-

active graphics in R, such as GGobi (2010) and iPlots (Urbanek et al. 2013), none of them seem to include support for beanplots. In any case, the use of R is not yet widespread in diachronic corpus linguistics owing to the steep learning curve.

As discussed in Chapters 10 and 12, the new implementation of the permutation testing method (*types*2, Suomela 2014) produces interactive type accumulation curves embedded on web pages. Figure 13.5 shows an example of a web page generated by the software. The text at the top describes the corpus and the kind of sample used in the permutation testing. The user has clicked on "ity" to view the results for the suffix *-ity*, on "sex-relcode" to view subcorpora based on gender and the relationship between the sender and the recipient of the letter, and on "types/running words" to view the results for types as a function of the number of running words. As the subcorpus "M, TC" (letters written by men to their close friends) looks interesting, the user has clicked on it, which highlights it on the graph and provides additional information on it at the bottom of the page. At 223 types, the productivity of *-ity* is significantly high in this subcorpus at $p = 0.000059$. According to the Corpus menu page (a separate web page, not shown in Figure 13.5), this is still significant after false discovery rate control at a rate of 0.01.

While this is definitely an improvement on the previous version (Suomela 2007), some functionality is still missing. Like most of the software used for data visualisation – including R, Matlab and Mondrian (Theus 2011) – *types*2 does not read in the text of the corpus, so the connection to the text is inevitably lost, and the complexity of the text is reduced to a few numerical measurements (cf. Siirtola et al. 2014). With large corpora, linking interactive visualisations directly to the full texts might take too long to execute, so a better solution might be to generate lists of links to the full texts with metadata displayed next to the links, or to link to the relevant concordance lines with a limited amount of context. Even this could be challenging because corpora come in so many different formats, so perhaps the visualisations should be integrated in a standard corpus-linguistic toolkit equipped to deal with many kinds of corpora. This would also make them accessible to everyone in the field, although more experienced users might find the options provided by any ready-made toolkit too limiting (cf. Anthony 2013).

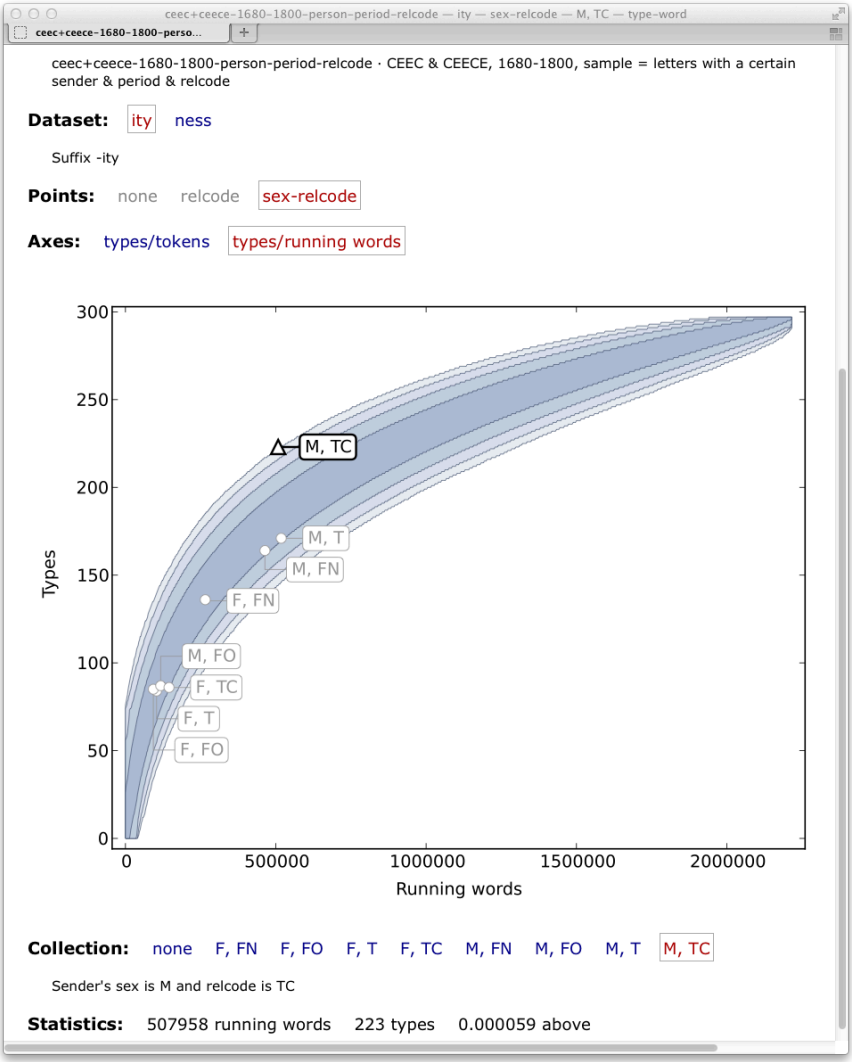

Figure 13.5. Web page generated by the *types*2 software (Suomela 2014)

14. Evaluation and explanation of results

This chapter discusses the linguistic results presented in Part II above, relating them to previous studies as well as providing suggestions for future research. The focus is on each suffix separately, beginning with *-ity*.

14.1. Variable and increasingly productive *-ity*

Even though the picture emerging of *-ity* in my studies is one of continuous growth in productivity, the shape of its type and hapax accumulation curves in the material used speaks for its relative lack of productivity compared to *-ness* (see 13.1.2 above). It would be of interest to analyse the shape of the curves in a moderately sized corpus representing a genre of Present-day English in which *-ity* is known to be highly productive, such as a specific subgenre of scientific writing, to see whether the bounds for hapax legomena would be any narrower there and whether the steep initial rise and later levelling out would be replaced with a more linear kind of growth. It is possible that the labelling function of *-ity* inevitably leads to a different shape compared to *-ness*, which is used in more ephemeral formations in syntactic recategorisation. In their interpolated growth curves of PDE newspaper data, Baayen and Renouf (1996: 91) do not see much of a difference between *-ness* and *-ity*, so it would also be of interest to replicate their study using actual growth curves and the permutation testing approach proposed in this dissertation.

It remains somewhat unclear whether the increase in the productivity of *-ity* in the 18th-century section of the CEEC corpora is linguistic or stylistic, or both. While it has been connected to a stylistic change in the letter genre along Biber and Finegan's (1997) dimensions of register variation, these were studied using the ARCHER corpus, which was not compiled according to the same principles as the CEEC. Future research could conduct a similar multidimensional analysis in the CEEC, or at least trace variation in noun and pronoun frequencies in a similar manner as was done in the earlier centuries of the CEEC in Chapter 8. This will become increasingly possible after we have finished standardising the spelling of the 18th-century section and after its subsequent part-of-speech tagging (see Säily 2013). Vartiainen et al. (2013) have already studied the frequencies of first- and second-person pronouns in the 17th- and 18th-century sections of the CEEC. In accordance with the shared style hypothesis presented in Chapter 10, they find considerable levelling of gender differences in the 18th century. However, there are statistically significant intra-gender differences according to the

relationship between the sender and recipient of the letter, suggesting that these should be paid close attention to in future research.

Another interesting issue is the relatively stable gender difference in the use of *-ity* suggesting that women tend to use it less productively than men. Gender has also proved significant in previous research, including the studies by Keune et al. (2006) and Keune (2012: Chapter 4) of present-day Dutch, which found that men tended to use affixes more productively than women. It remains to be seen whether these results will be replicated with more affixes and data sets. It is clear, however, that this is not something that is invariably true. In my studies, no gender difference was found in the use of *-ness* in most corpora, and the overall gender difference in the use of *-ity* disappeared in 18th-century correspondence. Furthermore, differences between individuals are great. As mentioned in Section 5.1.1 above, the measure of statistical significance could be supplemented with a measure of effect size to facilitate a more accurate assessment of the importance of these and other results.

When analysing relatively rare phenomena in historical corpora, the social categories of age and region are more challenging to study than are gender and social rank, owing to sparse metadata and the small amount of data per group. Nevertheless, a small study of the former two categories would be of interest as they have proved relevant in previous research (see 2.3.4.3 above). I do not expect to see regional variation in the productivity of *-ity* and *-ness* after the turbulent Middle English period, but the hypothesis should certainly be tested. While Chapter 6 found no significant regional variation in 17th-century correspondence, this could be due to lack of data, so it might be useful to combine some regions and perhaps extend the analysis to earlier centuries. The category of age might well be a factor in the increase in the productivity of *-ity*.

From a wider perspective, the variation and change in the productivity of *-ity* could be connected to the paradigm of near-synonymous nominal suffixes as a whole. It could be argued that after the relative chaos of Middle English and the *copia* of the 16th century (see 1.2 above), the system became more streamlined, enabling a steady increase in the productivity of *-ity*. While the productivity of many suffixes increased in the 16th century, this seems to have been productivity of an ephemeral sort: as noted by Nevalainen (1999a: 349), referring to Neuhaus's (1971) study of the *Shorter Oxford English Dictionary*, "the intensive period of neologising is followed by a corresponding increase in obsolete words".

14.2. Default suffix -*ness*?

Synchronic research has seen -*ness* as a default suffix, and there does indeed seem to be less sociolinguistic variation in its productivity than in that of -*ity* in my material. Less variation implies less change: I observed no changes in its productivity. This does not mean, however, that the productivity of -*ness* has never changed. Previous research has shown that it has been in competition with a range of other affixes and that its productivity has changed across base types as well as overall compared to other nominal suffixes in the paradigm (see Sections 2.4 and 3.2 above). As English began to be used in more and more functions, the increased need for new English words caused an increase in the productivity of -*ness* as well as many other suffixes. As noted in the previous section, the productivity seems to have been of an ephemeral kind in the 16th century, after which the situation was streamlined.

I have only studied -*ness* in two historical genres, personal correspondence and trial proceedings. It is therefore possible that its productivity increased during, e.g., the 17th century in other genres that I did not study. I have selected speech-related genres because speech is the primary medium of language and the origin of most changes; thus, it is the preferred object of study in sociolinguistics (Nevalainen and Raumolin-Brunberg 2003: 28). However, nominal suffixes are clearly less productive in speech than in writing. Here we again run into the question of linguistic versus stylistic change: if we wish to study change in the grammar of the speech community as a whole, we need to study either a corpus representing the language as a whole or a speech-related corpus, and I have chosen the latter option because it is the only option for which we have corpora that include sociolinguistic metadata. Moreover, speech-related corpora cover the widest range of language users in terms of social categories such as gender and social status.

In addition to variation in productivity, I have observed change in the semantics of both -*ness* and -*ity* between the 17th and 18th centuries. It seems that these suffixes can be put to multiple uses depending on the needs of the language users. While existing words in previously observed contexts serve as models, there is a great deal of semantic variation among them, and even rare senses can be brought to the foreground when the situation so demands. In 18th-century correspondence, the situation seems to more and more often demand an involved style of writing, so that middle- and upper-class writers recruit -*ness* and -*ity* words increasingly frequently to describe embodied attributes or traits. In -*ness*, this was already a prominent sense, which only became more so. It would be

interesting to conduct a similar study in, e.g., the *Old Bailey Corpus*, in which I would not expect to see a more involved style.

15. Implications for future research

This chapter concludes the dissertation by providing more directions for future research, first zooming in on *-ness* and *-ity*, then widening the focus to socio-linguistic variation in productivity and to corpus-linguistic methodology.

15.1. Zooming in: Structures and functions

The present work has studied *-ness* and *-ity* from the perspective of socio-linguistic variation in their morphological productivity. However, there is more to their story. I am interested in (changes in) the functions in which they are used as well as the kinds of grammatical constructions in which they appear.

We have already seen that the 'embodied attribute or trait' sense of *-ness* and *-ity* is associated with an involved style and with possessive constructions, as in *your kindness* or *the wetness of the weather* (Chapter 11 above). Essentially, both constructions are nominalised versions of the clause "x BE y" (*you are kind*, *the weather is wet*). The latter construction has been identified as typical of certain 18th-century authors fond of abstract diction (Bax 2005), some of whose letters are included in the CEECE. It would be worthwhile to analyse socio-linguistic variation and change in the extent to which instances of *-ness* and *-ity* occur in these constructions, taking into account linguistic factors such as the animacy of the head x.

Another area in need of further research concerns the functions of *-ness* and *-ity*, more specifically, sociolinguistic variation and change in the extent to which they are used for syntactic recategorisation. Recall that this function has been considered to be more typical of *-ness* than of *-ity* and that it has been associated with greater productivity as well as with the 'embodied attribute or trait' meaning (Section 2.4 above). However, there seem to have been no corpus-linguistic studies to ascertain whether this really is the case. The concept of recategorisation could be operationalised as the requirement that the base of the formation needs to occur before the formation in the text. I have already conducted a pilot study raising the issue of the functions and constructions in which *-ness* occurs among a few individuals in the CEECE.

15.2. Morphological productivity and corpus-linguistic methodology

The results of my work support the notion of sociolinguistic variation in morphological productivity. Each of the social categories studied – gender, social rank, and register in terms of participant relations – has turned out to be relevant,

gender being the most consistent factor. All of these should be taken into account in future studies. The relative importance of social rank and register in changes in productivity remains an issue for future research. Another question of interest is whether less productive affixes are more susceptible to sociolinguistic variation and change, as would seem to be the case based on my studies of *-ness* and *-ity* (see also Palmer 2009: 335).

This dissertation has presented a reliable and data-driven solution to the problem of comparing type and hapax frequencies across sociolinguistically defined subcorpora. The method significantly facilitates the study of variation and change in productivity. The amount of data permitting, future work could also use the method to more reliably compare frequencies of hybrid formations, which have been used as another measure of productivity in diachrony.

The present work has demonstrated the value of exploratory corpus analysis that uses robust statistics and helpful visualisations. Complemented by an intimate knowledge of the texts and the contexts in which they were produced, this approach enables even relatively small corpora to yield interesting and reliable results. It has become clear that when comparing word frequencies between corpora containing multiple texts, we should employ dispersion-aware tests of statistical significance. In particular, tests based on resampling have been recommended because they make the best use of the available data while entailing few background assumptions. Visualising the results using graphs that reveal the variability within (sub)corpora helps the analyst to discover outliers and to assess the importance of the results. To facilitate exploration, we need to develop more interactive visualisation tools that preserve the connection to the texts and metadata in the corpora used. This will require an increasing amount of multidisciplinary collaboration of the kind employed in this dissertation.

Bibliography

Adamson, Sylvia 1989. With double tongue: diglossia, stylistics and the teaching of English. Mick Short (ed.), *Reading, Analysing and Teaching Literature*, 204–240. London: Longman.

Adamson, Sylvia 1999. Literary language. Roger Lass (ed.), *The Cambridge History of the English Language, III: 1476–1776*, 539–653. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521264761.008.

Anderson, Karen 2000. Productivity in English nominal and adjectival derivation, 1100–2000. PhD dissertation, University of Western Australia.

Anshen, Frank and Mark Aronoff 1989. Morphological productivity, word frequency and the Oxford English Dictionary. Ralph W. Fasold and Deborah Schiffrin (eds.), *Language Change and Variation*, 197–202. Current Issues in Linguistic Theory 52. Amsterdam: John Benjamins.

Anthony, Laurence 2013. A critical look at software tools in corpus linguistics. *Linguistic Research* 30(2): 141–161.

Archer, Dawn 2012. Corpus annotation: a welcome addition or an interpretation too far? Jukka Tyrkkö, Matti Kilpiö, Terttu Nevalainen and Matti Rissanen (eds.), *Outposts of Historical Corpus Linguistics: From the Helsinki Corpus to a Proliferation of Resources*. Studies in Variation, Contacts and Change in English 10. Helsinki: VARIENG. http://www.helsinki.fi/varieng/series/volumes/10/archer/ (accessed 9 May 2014).

Archer, Dawn 2014. Exploring verbal aggression in English historical texts using USAS: the possibilities, the problems and potential solutions. Irma Taavitsainen, Andreas H. Jucker and Jukka Tuominen (eds.), *Diachronic Corpus Pragmatics*, 277–302. Pragmatics & beyond New Series 243. Amsterdam: John Benjamins. doi:10.1075/pbns.243.17arc.

Argamon, Shlomo, Moshe Koppel, Jonathan Fine and Anat Rachel Shimoni 2003. Gender, genre, and writing style in formal written texts. *Text* 23(3): 321–346.

Aronoff, Mark 1976. *Word Formation in Generative Grammar*. Linguistic Inquiry Monographs 1. Cambridge, Massachusetts: The MIT Press.

Aronoff, Mark and Frank Anshen 1998. Morphology and the lexicon: lexicalization and productivity. Andrew Spencer and Arnold M. Zwicky (eds.), *The Handbook of Morphology*, 237–247. Cambridge, Massachusetts: Blackwell Publishers.

Aronoff, Mark and Livio Gaeta 2003. Introduction. *Italian Journal of Linguistics* 15(1): 3–6. http://linguistica.sns.it/RdL/2003.html (accessed 9 May 2014).

Aronoff, Mark and Roger Schvaneveldt 1978. Testing morphological productivity. *Annals of the New York Academy of Sciences: Papers in Anthropology and Linguistics* 318: 106–114. doi:10.1111/j.1749-6632.1978.tb16357.x.

Baayen, R. H. 1992. Quantitative aspects of morphological productivity. Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1991*, 109–149. Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-011-2516-1_8.

Baayen, R. H. 1993. On frequency, transparency and productivity. Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1992*, 181–208. Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-017-3710-4_7.

Baayen, R. H. 1994. Derivational productivity and text typology. *Journal of Quantitative Linguistics* 1(1): 16–34. doi:10.1080/09296179408589996.

Baayen, R. H. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-010-0844-0.

Baayen, R. H. 2009. Corpus linguistics in morphology: morphological productivity. Anke Lüdeling and Merja Kytö (eds.), *Corpus Linguistics: An International Handbook*, 899–919. Berlin: Mouton de Gruyter. doi:10.1515/9783110213881.2.899.

Baayen, R. H. and Rochelle Lieber 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29(5): 801–843. doi:10.1515/ling.1991.29.5.801.

Baayen, R. H. and Anneke Neijt 1997. Productivity in context: a case study of a Dutch suffix. *Linguistics* 35(3): 565–587. doi:10.1515/ling.1997.35.3.565.

Baayen, R. H. and Antoinette Renouf 1996. Chronicling the *Times*: productive lexical innovations in an English newspaper. *Language* 72(1): 69–96. http://www.jstor.org/stable/416794 (accessed 9 May 2014).

Baeskow, Heike 2012. *-Ness* and *-ity*: phonological exponents of *n* or meaningful nominalizers of different adjectival domains? *Journal of English Linguistics* 40(1): 6–40. doi:10.1177/0075424211405156.

Baker, Paul 2010. Times may change, but we will always have money: diachronic variation in Recent British English. *Journal of English Linguistics* 39(1): 65–88. doi:10.1177/0075424210368368.

Barber, Charles 1976. *Early Modern English*. London: André Deutsch.

Baron, Alistair 2011. VARD 2. Computer program. http://www.comp.lancs.ac.uk/~barona/vard2/ (accessed 9 May 2014).

Baron, Alistair and Paul Rayson 2009. Automatic standardisation of texts containing spelling variation: how much training data do you need? Michaela Mahlberg, Victorina González-Díaz and Catherine Smith (eds.), *Proceedings of the Corpus Linguistics Conference: CL2009, University of Liverpool, UK, 20–23 July 2009*, article #314. http://ucrel.lancs.ac.uk/publications/cl2009/ (accessed 9 May 2014).

Bauer, Laurie 1983. *English Word-Formation*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Bauer, Laurie 2001. *Morphological Productivity*. Cambridge Studies in Linguistics 95. Cambridge: Cambridge University Press.

Bauer, Laurie, Rochelle Lieber and Ingo Plag 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.

Bax, Randy 2005. Traces of Johnson in the language of Fanny Burney. *International Journal of English Studies* 5(1): 159–181. http://revistas.um.es/ijes/article/view/47941 (accessed 9 May 2014).

Bell, Allan 1984. Language style as audience design. *Language in Society* 13(2): 145–204. doi:10.1017/S004740450001037X.

Benjamini, Yoav and Yosef Hochberg 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289–300. http://www.jstor.org/stable/10.2307/2346101 (accessed 9 May 2014).

Bentz, Christian, Douwe Kiela, Felix Hill and Paula Buttery forthcoming. Zipf's law and the grammar of languages: a quantitative study of Old and Modern English parallel texts. *Corpus Linguistics and Linguistic Theory*. doi:10.1515/cllt-2014-0009.

Bergs, Alexander 2012. The Uniformitarian Principle and the risk of anachronisms in language and social history. Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics*, 80–98. Blackwell Handbooks in Linguistics. Chichester: Wiley-Blackwell. doi:10.1002/9781118257227.ch5.

Bestgen, Yves 2014. Inadequacy of the chi-squared test to examine vocabulary differences between corpora. *Literary and Linguistic Computing* 29(2): 164–170. doi:10.1093/llc/fqt020.

Biber, Douglas 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511621024.

Biber, Douglas 2001. Dimensions of variation among 18th-century speech-based and written registers. Hans-Jürgen Diller and Manfred Görlach (eds.), *Towards a History of English as a History of Genres*, 89–109. Anglistische Forschungen 298. Heidelberg: Winter.

Biber, Douglas and Jená Burges 2000. Historical change in the language use of women and men: gender differences in dramatic dialogue. *Journal of English Linguistics* 28(1): 21–37. doi:10.1177/00754240022004857.

Biber, Douglas and Edward Finegan 1989. Drift and the evolution of English style: a history of three genres. *Language* 65(3): 487–517. http://www.jstor.org/stable/10.2307/415220 (accessed 9 May 2014).

Biber, Douglas and Edward Finegan 1997. Diachronic relations among speech-based and written registers in English. Terttu Nevalainen and Leena Kahlas-Tarkka (eds.), *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*, 253–275. Mémoires de la Société Néophilologique de Helsinki 52. Helsinki: Société Néophilologique.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Blythe, Richard A. and William Croft 2012. S-curves and the mechanisms of propagation in language change. *Language* 88(2): 269–304. doi:10.1353/lan.2012.0027.

BNC = The *British National Corpus*, version 3 (BNC XML edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. http://www.natcorp.ox.ac.uk (accessed 9 May 2014).

Bolinger, Dwight L. 1948. On defining the morpheme. *Word* 4: 18–23.

Booij, Geert 2007. Construction morphology and the lexicon. Fabio Montermini, Gilles Boyé and Nabil Hathout (eds.), *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*, 34–44. Somerville, Massachusetts: Cascadilla Proceedings Project. http://www.lingref.com/cpp/decemb/5/abstract1613.html (accessed 9 May 2014).

Booij, Geert 2010. *Construction Morphology*. Oxford: Oxford University Press.

Březina, Václav 2005. The development of the prefixes *un-* and *in-* in Early Modern English with special regard to the sociolinguistic background. MA thesis, Faculty of Arts, Charles University in Prague.

Březina, Václav 2013. BNC64: comparison of male and female speech. Data used in the BNC64 Search & Compare tool have been extracted from the British National Corpus, distributed by Oxford University Computing Services on behalf of the BNC Consortium. Used with permission. http://corpora.lancs.ac.uk/bnc64/ (accessed 9 May 2014).

Brinton, Laurel J. and Elizabeth Closs Traugott 2005. *Lexicalization and Language Change*. Research Surveys in Linguistics. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511615962.

Burnard, Lou (ed.) 2007. Reference guide for the British National Corpus (XML edition). Published for the British National Corpus Consortium by the Research Technologies Service at Oxford University Computing Services. http://www.natcorp.ox.ac.uk/docs/URG/ (accessed 9 May 2014).

Cameron, Deborah 2006. Gender. Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 733–739. 2nd edition. Oxford: Elsevier. doi:10.1016/B0-08-044854-2/01463-2.

Cameron, Deborah 2008. Issues of gender in modern English. Haruko Momma and Michael Matto (eds.), *A Companion to the History of the English Language*, 292–302. Chichester: Wiley-Blackwell. doi:10.1002/9781444302851.ch29.

Campbell, George 1776. *The Philosophy of Rhetoric […]. Volume 1*. London: printed for W. Strahan; and T. Cadell, in the Strand; and W. Creech at Edinburgh.

Cannadine, David 2000 [1998]. *Class in Britain*. London: Penguin Books.

Cantos, Pascual 2012. The use of linguistic corpora for the study of linguistic variation and change: types and computational applications. Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics*, 99–122. Blackwell Handbooks in Linguistics. Chichester: Wiley-Blackwell. doi:10.1002/9781118257227.ch6.

CEEC = *Corpus of Early English Correspondence*. 1998. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi and Minna Palander-Collin at the Department of Modern Languages, University of Helsinki. http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/ (accessed 9 May 2014).

CEECE = *Corpus of Early English Correspondence Extension*. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Samuli Kaislaniemi, Mikko Laitinen, Minna Nevala, Arja Nurmi, Minna Palander-Collin, Tanja Säily and Anni Sairio at the Department of Modern Languages, University of Helsinki. http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/ (accessed 9 May 2014).

CEECSU = *Corpus of Early English Correspondence Supplement*. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Samuli Kaislaniemi, Mikko Laitinen, Minna Nevala, Arja Nurmi, Minna Palander-Collin, Tanja Säily and Anni Sairio at the Department of Modern Languages, University of Helsinki. http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/ (accessed 9 May 2014).

Church, Kenneth W. and William A. Gale 1995. Poisson mixtures. *Natural Language Engineering* 1(2): 163–190. doi:10.1017/S1351324900000139.

Coolidge, Frederick L. 2013. *Statistics: A Gentle Introduction*. 3rd edition. Thousand Oaks, California: Sage.

CoRD = Corpus Resource Database. Research Unit for Variation, Contacts and Change in English, University of Helsinki. http://www.helsinki.fi/varieng/CoRD/ (accessed 9 May 2014).

Corpora of Early English Correspondence. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg et al. at the Department of Modern Languages, University of Helsinki. http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/ (accessed 9 May 2014).

Corpus of Early English Correspondence Sampler. 1998. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi and Minna Palander-Collin at the Department of Modern Languages, University of Helsinki. http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/ (accessed 9 May 2014).

Cowie, Claire 1999. Diachronic word-formation: a corpus-based study of derived nominalizations in the history of English. PhD dissertation, University of Cambridge.

Cowie, Claire 2003. "Uncommon terminations": proscription and morphological productivity. *Italian Journal of Linguistics* 15(1): 17–30. http://linguistica.sns.it/RdL/2003.html (accessed 9 May 2014).

Cowie, Claire and Christiane Dalton-Puffer 2002. Diachronic word-formation and studying changes in productivity over time: theoretical and methodological considerations. Javier E. Díaz Vera (ed.), *A Changing World of Words: Studies in English Historical Lexicography, Lexicology and Semantics*, 410–437. Costerus New Series 141. Amsterdam: Rodopi.

Culpeper, Jonathan and Merja Kytö 2010. *Early Modern English Dialogues: Spoken Interaction as Writing*. Studies in English Language. Cambridge: Cambridge University Press.

Culpeper, Jonathan and Minna Nevala 2012. Sociocultural processes and the history of English. Terttu Nevalainen and Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, 365–391. Oxford Handbooks in Linguistics. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199922765.013.0032.

Culy, Chris, Verena Lyding and Henrik Dittmann 2011. Structured Parallel Coordinates: a visualization for analyzing structured language data. María Luisa Carrió Pastor and Miguel Ángel Candel Mora (eds.), *CILC-11: Proceedings of the 3rd International Conference on Corpus Linguistics, Valencia, Spain, 6–9 April 2011*, 525–533. València: Editorial Universitat Politècnica de València. http://www.editorial.upv.es/publicacion/6032 (accessed 9 May 2014).

Dalton-Puffer, Christiane 1992. The status of word formation in Middle English: approaching the question. Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen and Irma Taavitsainen (eds.), *History of Englishes: New Methods and Interpretations in Historical Linguistics*, 465–482. Berlin: Mouton de Gruyter. doi:10.1515/9783110877007.465.

Dalton-Puffer, Christiane 1996. *The French Influence on Middle English Morphology: A Corpus-Based Study of Derivation*. Topics in English Linguistics 20. Berlin: Mouton de Gruyter.

Davies, Mark 2010. The Corpus of Historical American English, Google Books, and our new Google Books interface. http://corpus.byu.edu/coha/compare-googleBooks.asp (accessed 9 May 2014).

Diller, Hans-Jürgen, Hendrik De Smet and Jukka Tyrkkö 2010. A European database of descriptors of English electronic texts. *The European English Messenger* 19(2): 29–35. http://www.essenglish.org/mess/mestoc192.html (accessed 9 May 2014).

Dunning, Ted 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19(1): 61–74. http://dl.acm.org/citation.cfm?id=972454 (accessed 9 May 2014).

Durrell, Martin, Astrid Ensslin and Paul Bennett 2007. GerManC: a historical corpus of German 1650–1800. *Sprache und Datenverarbeitung* 31: 71–80.

Dwass, Meyer 1957. Modified Randomization Tests for Nonparametric Hypotheses. *The Annals of Mathematical Statistics* 28(1): 181–187. http://projecteuclid.org/euclid.aoms/1177707045 (accessed 9 May 2014).

EEBO = *Early English Books Online*. ProQuest. http://eebo.chadwyck.com (accessed 9 May 2014).

Efron, Bradley and Robert J. Tibshirani 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

English Civil War. Wikipedia, the Free Encyclopedia. http://en.wikipedia.org/wiki/English_Civil_War (accessed 9 May 2014).

Evans, Mel 2012. A sociolinguistics of Early Modern spelling? An account of Queen Elizabeth I's correspondence. Jukka Tyrkkö, Matti Kilpiö, Terttu Nevalainen and Matti Rissanen (eds.), *Outposts of Historical Corpus Linguistics: From the Helsinki Corpus to a Proliferation of Resources*. Studies in Variation, Contacts and Change in English 10. Helsinki: VARIENG. http://www.helsinki.fi/varieng/series/volumes/10/evans/ (accessed 9 May 2014).

Evert, Stefan 2006. How random is a corpus? The library metaphor. *Zeitschrift für Anglistik und Amerikanistik* 54(2): 177–190. doi:10.1515/zaa-2006-0208.

Evert, Stefan 2009. Corpora and collocations. Anke Lüdeling and Merja Kytö (eds.), *Corpus Linguistics: An International Handbook*, vol. 2, 1212–1248. Handbooks of Linguistics and Communication Science 29/2. Mouton de Gruyter. doi:10.1515/9783110213881.2.1212.

Evert, Stefan and Marco Baroni 2005. Testing the extrapolation quality of word frequency models. Pernilla Danielsson and Martijn Wagenmakers (eds.), *Proceedings of Corpus Linguistics 2005*. The Corpus Linguistics Conference Series 1. http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx (accessed 9 May 2014).

Evert, Stefan and Marco Baroni 2007. zipfR: word frequency distributions in R. Sophia Ananiadou (ed.), *Proceedings of the ACL 2007 Demo and Poster Sessions*, 29–32. Stroudsburg, Pennsylvania: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1557780 (accessed 9 May 2014).

Fenning, Daniel 1771. *A New Grammar of the English Language; or an Easy Introduction to the Art of Speaking and Writing English with Propriety and Correctness: […]*. London: printed for S. Crowder.

Finegan, Edward and Douglas Biber 2001. Register variation and social dialect variation: the Register Axiom. Penelope Eckert and John R. Rickford (eds.), *Style and Sociolinguistic Variation*, 235–267. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511613258.015.

Fisher, R. A. 1922. On the interpretation of χ^2 from contingency tables, and the calculation of *p*. *Journal of the Royal Statistical Society* 85(1): 87–94. doi:10.2307/2340521.

Fitzmaurice, Susan 2002. *The Familiar Letter in Early Modern English: A Pragmatic Approach*. Pragmatics & beyond New Series 95. Amsterdam: John Benjamins.

Fitzmaurice, Susan 2012. Social factors and language change in eighteenth-century England: the case of multiple negation. *Neuphilologische Mitteilungen* 113(3): 293–321.

Fletcher, Anthony 1996. *Gender, Sex and Subordination in England 1500–1800*. New Haven: Yale University Press.

Foyster, Elizabeth A. 1999. *Manhood in Early Modern England: Honour, Sex and Marriage*. Women and Men in History. London: Longman.

Frauenfelder, Uli H. and Robert Schreuder 1992. Constraining psycholinguistic models of morphological processing and representation: the role of productivity. Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1991*, 165–183. Dordrecht: Kluwer Academic Publishers. doi:10.1007/978-94-011-2516-1_10.

Friendly, Michael and Daniel Denis 2005. The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences* 41(2): 103–130. doi:10.1002/jhbs.20078.

Gabrielatos, Costas and Anna Marchi 2012. Keyness: appropriate metrics and practical issues. *CADS International Conference 2012. Corpus-Assisted Discourse Studies: More than the Sum of Discourse Analysis and Computing?, 13–14 September, University of Bologna, Italy*. Conference presentation. http://repository.edgehill.ac.uk/id/eprint/4196 (accessed 9 May 2014).

Gaeta, Livio and Davide Ricca 2006. Productivity in Italian word formation: a variable-corpus approach. *Linguistics* 44(1): 57–89.

Gardner, Anne 2013. Derivation in Middle English: regional and text type variation. PhD dissertation, University of Zurich.

Garey, Michael R. and David S. Johnson 2003 [1979]. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: W. H. Freeman.

Garretson, Gregory 2008. Desiderata for linguistic software design. *International Journal of English Studies* 8(1): 67–94. http://revistas.um.es/ijes/article/view/49101 (accessed 9 May 2014).

Gesmann, Markus and Diego de Castillo 2011. Using the Google Visualisation API with R. *The R Journal* 3(2): 40–44. http://journal.r-project.org/archive/2011-2/ (accessed 9 May 2014).

GGobi data visualization system. 2010. Computer program. The GGobi Foundation. http://www.ggobi.org (accessed 9 May 2014).

Good, Phillip 2005. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. 3rd edition. Springer Series in Statistics. Berlin: Springer. doi:10.1007/b138696.

Görlach, Manfred 2001. *Eighteenth-Century English*. Sprachwissenschaftliche Studienbücher. Heidelberg: Winter.

Gotelli, Nicholas J. and Robert K. Colwell 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4(4): 379–391. doi:10.1046/j.1461-0248.2001.00230.x.

Greenwood, James 1722. *An Essay towards a Practical English Grammar, […]*. 2nd edition. London: printed for John Clark.

Gries, Stefan Th. 2005. Null-hypothesis significance testing of word frequencies: a follow-up on Kilgarriff. *Corpus Linguistics and Linguistic Theory* 1(2): 277–294. doi:10.1515/cllt.2005.1.2.277.

Gries, Stefan Th. 2006. Some proposals towards a more rigorous corpus linguistics. *Zeitschrift für Anglistik und Amerikanistik* 54(2): 191–202. doi:10.1515/zaa-2006-0209.

Gries, Stefan Th. 2006b. Exploring variability within and between corpora: some methodological considerations. *Corpora* 1(2): 109–151. doi:10.3366/cor.2006.1.2.109.

Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13(4): 403–437. doi:10.1075/ijcl.13.4.02gri.

Gries, Stefan Th. and Martin Hilpert 2008. The identification of stages in diachronic data: variability-based neighbour clustering. *Corpora* 3(1): 59–81. doi:10.3366/E1749503208000075.

Halliday, M. A. K. 1978. *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

Hardie, Andrew 2007. Part-of-speech ratios in English corpora. *International Journal of Corpus Linguistics* 12(1): 55–81. doi:10.1075/ijcl.12.1.05har.

Hartigan, J. A. and Beat Kleiner 1984. A mosaic of television ratings. *The American Statistician* 38(1): 32–35. http://www.jstor.org/stable/2683556 (accessed 9 May 2014).

Hay, Douglas and Nicholas Rogers 1997. *Eighteenth-Century English Society: Shuttles and Swords*. Oxford: Oxford University Press.

Hay, Jennifer 2001. Lexical frequency in morphology: is everything relative? *Linguistics* 39(6): 1041–1070.

Hay, Jennifer and R. H. Baayen 2003. Phonotactics, parsing and productivity. *Italian Journal of Linguistics* 15(1): 99–130. http://linguistica.sns.it/RdL/2003.html (accessed 9 May 2014).

Hay, Jennifer and Ingo Plag 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory* 22(3): 565–596. doi:10.1023/B:NALA.0000027679.63308.89.

HC = *The Helsinki Corpus of English Texts*. 1991. 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). http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/ (accessed 9 May 2014).

Hilpert, Martin 2011. Dynamic visualizations of language change: motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics* 16(4): 435–461. doi:10.1075/ijcl.16.4.01hil.

Hilpert, Martin 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Studies in English Language. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139004206.

Hilpert, Martin and Stefan Th. Gries forthcoming. Quantitative approaches to diachronic corpus linguistics. Merja Kytö and Päivi Pahta (eds.), *The Cambridge Handbook of English Historical Linguistics*. Cambridge: Cambridge University Press.

Hinneburg, Alexander, Heikki Mannila, Samuli Kaislaniemi, Terttu Nevalainen and Helena Raumolin-Brunberg 2007. How to handle small samples: bootstrap and Bayesian methods in the analysis of linguistic change. *Literary and Linguistic Computing* 22(2): 137–150. doi:10.1093/llc/fqm006.

Hlava, Magdalena 2005. Aspects of morphological productivity: a corpus-based study of foreign and native prefixes in Early Modern English. MA thesis, Department of English, University of Vienna.

Hoffmann, Sebastian 2004. Using the OED quotations database as a corpus – a linguistic appraisal. *ICAME Journal* 28: 17–30. http://icame.uib.no/ij28/ (accessed 9 May 2014).

Hoffmann, Sebastian, Stefan Evert, Nicholas Smith, David Lee and Ylva Berglund Prytz 2008. *Corpus Linguistics with BNCweb – a Practical Guide*. English Corpus Linguistics 6. Frankfurt am Main: Peter Lang.

Hofland, Knut and Stig Johansson 1982. *Word Frequencies in British and American English*. Bergen: Norwegian Computing Centre for the Humanities.

Holmes, Janet 1998. Women's talk: the question of sociolinguistic universals. Jennifer Coates (ed.), *Language and Gender: A Reader*, 461–483. Malden, Massachusetts: Blackwell.

Holmes, Janet 1999 [1997]. Setting new standards: sound changes and gender in New Zealand English. *English World-Wide* 18(1): 107–142. Reprinted in *Cuadernos de Filología Inglesa* 8(1): 147–175. http://dialnet.unirioja.es/servlet/articulo?codigo=112485 (accessed 9 May 2014).

Holmes, Janet and Miriam Meyerhoff 2003. Different voices, different views: an introduction to current research in language and gender. Janet Holmes and Miriam Meyerhoff (eds.), *The Handbook of Language and Gender*, 1–17. Oxford: Blackwell.

Hope, Jonathan and Michael Witmore 2010. The Hundredth Psalm to the tune of "Green Sleeves": digital approaches to Shakespeare's language of genre. *Shakespeare Quarterly* 61(3): 357–390. doi:10.1353/shq.2010.0002.

HT = Kay, Christian, Jane Roberts, Michael Samuels and Irené Wotherspoon (eds.) 2009. *Historical Thesaurus of the Oxford English Dictionary*. OED Online. http://www.oed.com/thesaurus (accessed 9 May 2014).

Huber, Magnus 2007. The Old Bailey Proceedings, 1674–1834: evaluating and annotating a corpus of 18th- and 19th-century spoken English. Anneli Meurman-Solin and Arja Nurmi (eds.), *Annotating Variation and Change*. Studies in Variation, Contacts and Change in English 1. Helsinki: VARIENG. http://www.helsinki.fi/varieng/series/volumes/01/huber/ (accessed 9 May 2014).

Huddleston, Rodney and Geoffrey K. Pullum 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hudson, Richard A. 1994. About 37% of word-tokens are nouns. *Language* 70(2): 331–339. http://www.jstor.org/stable/10.2307/415831 (accessed 9 May 2014).

Hudson, Richard A. 1996. *Sociolinguistics*. 2nd edition. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166843.

Hundt, Marianne and Geoffrey Leech 2012. "Small is beautiful": on the value of standard reference corpora for observing recent grammatical change. Terttu Nevalainen and Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, 175–188. Oxford Handbooks in Linguistics. Oxford: Oxford University Press.

Hundt, Marianne and Christian Mair 1999. 'Agile' and 'uptight' genres: the corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2): 221–242. doi:10.1075/ijcl.4.2.02hun.

James, Deborah 1996. Women, men and prestige speech forms: a critical review. Victoria L. Bergvall, Janet M. Bing and Alice F. Freed (eds.), *Rethinking Language and Gender Research: Theory and Practice*, 98–125. London: Longman.

Kampstra, Peter 2008. Beanplot: a boxplot alternative for visual comparison of distributions. *Journal of Statistical Software* 28: Code Snippet 1. http://www.jstatsoft.org/v28/c01/ (accessed 9 May 2014).

Kastovsky, Dieter 1986. The problem of productivity in word formation. *Linguistics* 24(3): 585–600. doi:10.1515/ling.1986.24.3.585.

Katz, Slava M. 1996. Distribution of content words and phrases in text and language modelling. *Natural Language Engineering* 2(1): 15–59. doi:10.1017/S1351324996001246.

Kaufer, David, Cheryl Geisler, Pantelis Vlachos and Suguru Ishizaki 2006. Mining textual knowledge for writing education and research: the DocuScope project. Luuk van Waes, Mariëlle Leijten and Christophe M. Neuwirth (eds.), *Writing and Digital Media*, 115–129. Studies in Writing 17. Amsterdam: Elsevier. doi:10.1108/S1572-6304(2006)0000017011.

Kendall, Tyler 2011. Corpora from a sociolinguistic perspective. *Revista Brasileira de Linguística Aplicada* 11(2): 361–389. http://ref.scielo.org/3tmp99 (accessed 9 May 2014).

Keune, Karen 2012. Explaining register and sociolinguistic variation in the lexicon: corpus studies on Dutch. PhD dissertation, Radboud University Nijmegen. http://www.lotpublications.nl/publish/articles/004498/bookpart.pdf (accessed 9 May 2014).

Keune, Karen, Roeland van Hout and R. H. Baayen 2006. Socio-geographic variation in morphological productivity in spoken Dutch: a comparison of statistical techniques. Jean-Marie Viprey and Lexicometrica (eds.), *Actes de JADT 2006 : 8es journées internationales d'analyse statistique des données textuelles*, 571–581. Besançon: Université de Franche-Comté. http://lexicometrica.univ-paris3.fr/jadt/jadt2006/tocJADT2006.htm (accessed 9 May 2014).

Kilgarriff, Adam 1996. Which words are particularly characteristic of a text? A survey of statistical approaches. *Proceedings of the AISB Workshop on Language Engineering for Document Analysis and Recognition*, 33–40. Sussex University. http://www.kilgarriff.co.uk/Publications/1996-K-AISB.pdf (accessed 9 May 2014).

Kilgarriff, Adam 2001. Comparing corpora. *International Journal of Corpus Linguistics* 6(1): 97–133. doi:10.1075/ijcl.6.1.05kil.

Kilgarriff, Adam 2005. Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory* 1(2): 263–275. doi:10.1515/cllt.2005.1.2.263.

Kirkby, John 1746. *A New English Grammar*. London: for R. Manby & H. S. Coy. (Facs. ed. Menston: Scolar, 1971, EL 297).

Körtvélyessy, Lívia 2009. Productivity and creativity in word-formation: a sociolinguistics perspective. *Onomasiology Online* 10: 1–22. http://www1.ku-eichstaett.de/SLF/EngluVglSW/OnOn10.htm (accessed 9 May 2014).

Körtvélyessy, Lívia and Pavol Štekauer forthcoming. Derivation in a social context. Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*. Oxford Handbooks in Linguistics. Oxford: Oxford University Press.

Krantz, David H. 1999. The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association* 94: 1372–1381. http://www.jstor.org/stable/2669949 (accessed 9 May 2014).

Kytö, Merja 1991. *Variation and Diachrony, with Early American English in Focus: Studies on can/may and shall/will*. Bamberger Beiträge zur Englischen Sprachwissenschaft 28. Frankfurt am Main: Peter Lang.

Kytö, Merja and Atro Voutilainen 1998. Backdating the English Constraint Grammar Parser for the analysis of English historical texts. Richard M. Hogg and Linda van Bergen (eds.), *Historical Linguistics 1995, Volume 2: Germanic Linguistics. Selected Papers from the 12th International Conference on Historical Linguistics, Manchester, August 1995*, 149–166. Current Issues in Linguistic Theory 162. Amsterdam: John Benjamins.

Labov, William 1978 [1972]. *Sociolinguistic Patterns*. Oxford: Basil Blackwell.

Labov, William 1982. Building on empirical foundations. Winfred P. Lehmann and Yakov Malkiel (eds.), *Perspectives on Historical Linguistics: Papers from a Conference Held at the Meeting of the Language Theory Division, Modern Language Assn., San Francisco, 27–30 December 1979*, 17–92. Current Issues in Linguistic Theory 24. Amsterdam: John Benjamins.

Labov, William 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2(2): 205–254. doi:10.1017/S0954394500000338.

Labov, William 1994. *Principles of Linguistic Change, Volume 1: Internal Factors*. Language in Society 20. Oxford: Blackwell.

Labov, William 2001. *Principles of Linguistic Change, Volume 2: Social Factors*. Language in Society 29. Malden, Massachusetts: Blackwell.

Laitinen, Mikko 2006. Expressing human indefiniteness in English: typology and markedness of pronouns. Terttu Nevalainen, Juhani Klemola and Mikko Laitinen (eds.), *Types of Variation: Diachronic, Dialectal and Typological Interfaces*, 203–239. Studies in Language Companion Series 76. Amsterdam: John Benjamins.

Laslett, Peter 1965. *The World We Have Lost*. New York: Charles Scribner's Sons.

Lee, David Y. W. 2001. Genres, registers, text types, domains, and styles: clarifying the concepts and nevigating a path through the BNC jungle. *Language Learning & Technology* 5(3): 37–72. http://ro.uow.edu.au/artspapers/598/ (accessed 9 May 2014).

Leech, Geoffrey 2011. The modals ARE declining: reply to Neil Millar's "Modal verbs in TIME: frequency changes 1923–2006", International Journal of Corpus Linguistics 14:2 (2009), 191–220. *International Journal of Corpus Linguistics* 16(4): 547–564. doi:10.1075/ijcl.16.4.05lee.

Leech, Geoffrey and Roger Fallon 1992. Computer corpora – what do they tell us about culture? *ICAME Journal* 16: 29–50. http://icame.uib.no/journal.html (accessed 9 May 2014).

Leech, Geoffrey and Nicholas Smith 2005. Extending the possibilities of corpus-based research on English in the twentieth century: a prequel to LOB and FLOB. *ICAME Journal* 29: 83–98. http://icame.uib.no/ij29/ (accessed 9 May 2014).

Lijffijt, Jefrey 2012. Bootstrap test for R and Matlab. Computer program. http://users.ics.aalto.fi/lijffijt/bootstraptest/ (accessed 9 May 2014).

Lijffijt, Jefrey 2013. Computational methods for comparison and exploration of event sequences. PhD dissertation, Department of Information and Computer Science, Aalto University. http://urn.fi/URN:ISBN:978-952-60-5475-9 (accessed 9 May 2014).

Lijffijt, Jefrey and Stefan Th. Gries 2012. Correction to Stefan Th. Gries' "Dispersions and adjusted frequencies in corpora", International Journal of Corpus Linguistics 13:4

(2008), 403–437. *International Journal of Corpus Linguistics* 17(1): 147–149. doi:10.1075/ijcl.17.1.08lij.

Lijffijt, Jefrey, Terttu Nevalainen, Tanja Säily, Panagiotis Papapetrou, Kai Puolamäki and Heikki Mannila forthcoming. Significance testing of word frequencies in corpora. *Literary and Linguistic Computing*.

Lijffijt, Jefrey, Panagiotis Papapetrou, Kai Puolamäki and Heikki Mannila 2011. Analyzing word frequencies in large text corpora using inter-arrival times and bootstrapping. Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba and Michalis Vazirgiannis (eds.), *Proceedings of ECML-PKDD 2011 – Part II*, 341–357. Berlin: Springer. doi:10.1007/978-3-642-23783-6_22.

Lijffijt, Jefrey, Tanja Säily and Terttu Nevalainen 2012. CEECing the baseline: lexical stability and significant change in a historical corpus. Jukka Tyrkkö, Matti Kilpiö, Terttu Nevalainen and Matti Rissanen (eds.), *Outposts of Historical Corpus Linguistics: From the Helsinki Corpus to a Proliferation of Resources*. Studies in Variation, Contacts and Change in English 10. Helsinki: VARIENG. http://www.helsinki.fi/varieng/series/volumes/10/lijffijt_saily_nevalainen/ (accessed 9 May 2014).

Lindsay, Mark 2012. Rival suffixes: synonymy, competition, and the emergence of productivity. Angela Ralli, Geert Booij, Sergio Scalise and Athanasios Karasimos (eds.), *Morphology and the Architecture of Grammar: On-Line Proceedings of the 8th Mediterranean Morphology Meeting (MMM8)*, 192–203. Rio: University of Patras. http://lmgd.philology.upatras.gr/el/research/downloads/MMM8_Proceedings.pdf#page=192 (accessed 9 May 2014).

Lindsay, Mark and Mark Aronoff 2013. Natural selection in self-organizing morphological systems. Nabil Hathout, Fabio Montermini and Jesse Tseng (eds.), *Morphology in Toulouse: Selected Proceedings of Décembrettes 7*. LINCOM Studies in Theoretical Linguistics 51. Munich: Lincom.

Lloyd, Cynthia 2011. *Semantics and Word Formation: The Semantic Development of Five French Suffixes in Middle English*. Studies in Historical Linguistics 6. Bern: Peter Lang.

Lupica Spagnolo, Marta 2013. Morphologische Produktivität in deutschsprachigen Texten nicht nativer Autoren: eine korpuslinguistische Analyse. *Zeitschrift für germanistische Linguistik* 41(3): 339–376. doi:10.1515/zgl-2013-0021.

Lyding, Verena, Stefania Degaetano-Ortlieb, Ekaterina Lapshinova-Koltunski, Henrik Dittmann and Chris Culy 2012. Visualising linguistic evolution in academic discourse. Miriam Butt, Sheelagh Carpendale, Gerald Penn, Jelena Prokić and Michael Cysouw (eds.), *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, 44–48. Stroudsburg, Pennsylvania: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=2388662 (accessed 9 May 2014).

Mair, Christian, Marianne Hundt, Geoffrey Leech and Nicholas Smith 2002. Short term diachronic shifts in part-of-speech frequencies: a comparison of the tagged LOB and F-LOB corpora. *International Journal of Corpus Linguistics* 7(2): 245–264. doi:10.1075/ijcl.7.2.05mai.

Malmgren, R. Dean, Daniel B. Stouffer, Andriana S. L. O. Campanharo and Luís A. Nunes Amaral 2009. On universality in human correspondence activity. *Science* 325(5948): 1696–1700. doi:10.1126/science.1174562.

Mann, H. B. and D. R. Whitney 1947. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* 18(1): 50–60. http://projecteuclid.org/euclid.aoms/1177730491 (accessed 9 May 2014).

Mannila, Heikki, Terttu Nevalainen and Helena Raumolin-Brunberg 2013. Quantifying variation and estimating the effects of sample size on the frequencies of linguistic variables. Manfred Krug and Julia Schlüter (eds.), *Research Methods in Language Variation and Change*, 337–360. Cambridge: Cambridge University Press.

Marchand, Hans 1969. *The Categories and Types of Present-Day English Word-Formation: A Synchronic-Diachronic Approach*. 2nd edition. Munich: C. H. Beck'sche Verlagsbuchhandlung.

Marle, Jaap van 1986. The domain hypothesis: the study of rival morphological processes. *Linguistics* 24(3): 601–627. doi:10.1515/ling.1986.24.3.601.

MATLAB – the language of technical computing. 2014. Natick, Massachusetts: The MathWorks. http://www.mathworks.com/products/matlab/ (accessed 9 May 2014).

McIntosh, Carey 1986. *Common and Courtly Language: The Stylistics of Social Class in 18th-Century British Literature*. Philadelphia: University of Pennsylvania Press.

McIntosh, Carey 2008. British English in the long eighteenth century (1660–1830). Haruko Momma and Michael Matto (eds.), *A Companion to the History of the English Language*, 228–234. Chichester: Wiley-Blackwell. doi:10.1002/9781444302851.ch22.

MED = *Middle English Dictionary*. 2001. Electronic version. http://quod.lib.umich.edu/m/med/ (accessed 9 May 2014).

Mendelson, Sara and Patricia Crawford 1998. *Women in Early Modern England, 1550–1720*. Oxford: Clarendon Press.

Metcalfe, Lister 1771. *The rudiments of the English tongue; or, the principles of English grammar,* […]. 2nd edition. Newcastle: printed by T. Saint, for J. Wilkie, London.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 331(6014): 176–182. doi:10.1126/science.1199644.

Milroy, James 1992. *Linguistic Variation and Change: On the Historical Sociolinguistics of English*. Language in Society 19. Oxford: Blackwell.

Milroy, Lesley 1999. Women as innovators and norm-creators: the sociolinguistics of dialect leveling in a northern English city. Suzanne Wertheim, Ashlee C. Bailey and Monica Corston-Oliver (eds.), *Engendering Communication: Proceedings of the Fifth Berkeley Women and Language Conference*, 361–376. Berkeley, California: BWLG.

Milroy, Lesley and Matthew Gordon 2003. *Sociolinguistics: Method and Interpretation*. Language in Society 34. Malden, Massachusetts: Blackwell.

Mitzenmacher, Michael and Eli Upfal 2005. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge: Cambridge University Press.

Mukherjee, Joybrato 2006. Corpus linguistics and English reference grammars. Antoinette Renouf and Andrew Kehoe (eds.), *The Changing Face of Corpus Linguistics: Papers from the 24th International Conference on English Language Research on Computerized Corpora (ICAME 24)*, 337–354. Language and Computers: Studies in Practical Linguistics 55. Amsterdam: Rodopi.

Myers, Sylvia Harcstark 1990. *The Bluestocking Circle: Women, Friendship, and the Life of the Mind in Eighteenth-Century England*. Oxford: Clarendon Press.

National Readership Survey 2014. Lifestyle data. http://www.nrs.co.uk/lifestyle-data/ (accessed 9 May 2014).

Nesselhauf, Nadja 2007. The spread of the progressive and its 'future' use. *English Language and Linguistics* 11(1): 191–207. doi:10.1017/S1360674306002152.

Neuhaus, H. J. 1971. Towards a diachronic analysis of vocabulary. *Cahiers de lexicologie* 18: 29–42.

Nevala, Minna 2007. Inside and out: forms of address in seventeenth- and eighteenth-century letters. Terttu Nevalainen and Sanna-Kaisa Tanskanen (eds.), *Letter Writing*, 89–113. Benjamins Current Topics 1. Amsterdam: John Benjamins. doi:10.1075/bct.1.07nev.

Nevala, Minna and Minna Palander-Collin 2005. Letters and letter writing: introduction. *European Journal of English Studies* 9(1): 1–7. doi:10.1080/13825570500067903.

Nevalainen, Terttu 1996. Social stratification. Terttu Nevalainen and Helena Raumolin-Brunberg (eds.), *Sociolinguistics and Language History: Studies Based on the Corpus of Early English Correspondence*, 57–76. Language and Computers: Studies in Practical Linguistics 15. Amsterdam: Rodopi.

Nevalainen, Terttu 1997. The processes of adverb derivation in Late Middle and Early Modern English. Matti Rissanen, Merja Kytö and Kirsi Heikkonen (eds.), *Grammaticalization at Work: Studies of Long-Term Developments in English*, 145–190. Topics in English Linguistics 24. Berlin: Mouton de Gruyter. doi:10.1515/9783110810745.145.

Nevalainen, Terttu 1999a. Early Modern English lexis and semantics. Roger Lass (ed.), *The Cambridge History of the English Language, III: 1476–1776*, 332–458. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521264761.006.

Nevalainen, Terttu 1999b. Making the best use of 'bad' data: evidence for sociolinguistic variation in Early Modern English. *Neuphilologische Mitteilungen* 100(4): 499–533.

Nevalainen, Terttu 2000. Processes of supralocalisation and the rise of Standard English in the Early Modern period. Ricardo Bermúdez-Otero, David Denison, Richard M. Hogg and C. B. McCully (eds.), *Generative Theory and Corpus Studies: A Dialogue from 10 ICEHL*, 329–371. Topics in English Linguistics 31. Berlin: Mouton de Gruyter. doi:10.1515/9783110814699.329.

Nevalainen, Terttu 2002. Language and woman's place in earlier English. *Journal of English Linguistics* 30(2): 181–199. doi:10.1177/007242030002006.

Nevalainen, Terttu 2006. Synchronic and diachronic variation. Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 356–363. 2nd edition. Oxford: Elsevier. doi:10.1016/B0-08-044854-2/01521-2.

Nevalainen, Terttu 2008. Variation in written English: grammar change or a shift in style? Susan Kermas and Maurizio Gotti (eds.), *Socially-Conditioned Language Change: Diachronic and Synchronic Insights*, 31–51. Lecce: Edizioni del Grifo.

Nevalainen, Terttu 2009. Grasshoppers and blind beetles: caregiver language in Early Modern English correspondence. Arja Nurmi, Minna Nevala and Minna Palander-Collin (eds.), *The Language of Daily Life in England (1400–1800)*, 137–164. Pragmatics & beyond New Series 183. Amsterdam: John Benjamins.

Nevalainen, Terttu 2012. Reconstructing syntactic continuity and change in Early Modern English regional dialects: the case of *who*. David Denison, Ricardo Bermúdez-Otero,

Chris McCully and Emma Moore (eds.), *Analysing Older English*, 159–184. Studies in English Language. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139022170.015.

Nevalainen, Terttu 2013. English historical corpora in transition: from new tools to legacy corpora? Paul Bennett, Martin Durrell, Silke Scheible and Richard J. Whitt (eds.), *New Methods in Historical Corpora*. Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 3. Tübingen: Narr.

Nevalainen, Terttu, Minna Palander-Collin and Tanja Säily (eds.) forthcoming. *Change in 18th-Century English: New Approaches to Historical Sociolinguistics*.

Nevalainen, Terttu and Helena Raumolin-Brunberg 2003. *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. Longman Linguistics Library. London: Pearson Education.

Nevalainen, Terttu and Helena Raumolin-Brunberg 2012. Historical sociolinguistics: origins, motivations, and paradigms. Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics*, 22–40. Blackwell Handbooks in Linguistics. Chichester: Wiley-Blackwell. doi:10.1002/9781118257227.ch2.

Nevalainen, Terttu and Ingrid Tieken-Boon van Ostade 2006. Standardisation. Richard Hogg and David Denison (eds.), *A History of the English Language*, 271–311. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511791154.006.

Nevalainen, Terttu and Heli Tissari 2010. Contextualising eighteenth-century politeness: social distinction and metaphorical levelling. Raymond Hickey (ed.), *Eighteenth-Century English: Ideology and Change*, 133–158. Studies in English Language. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511781643.009.

Nurmi, Arja 1999. *A Social History of Periphrastic DO*. Mémoires de la Société Néophilologique de Helsinki 56. Helsinki: Société Néophilologique.

Nurmi, Arja, Minna Nevala and Minna Palander-Collin (eds.) 2009. *The Language of Daily Life in England (1400–1800)*. Pragmatics & beyond New Series 183. Amsterdam: John Benjamins.

Nurmi, Arja and Minna Palander-Collin 2008. Letters as a text type: interaction in writing. Marina Dossena and Ingrid Tieken-Boon van Ostade (eds.), *Studies in Late Modern English Correspondence: Methodology and Data*, 21–49. Linguistic Insights: Studies in Language and Communication 76. Bern: Peter Lang.

Oakes, Michael P. and Malcolm Farrow 2007. Use of the chi-squared test to examine vocabulary differences in English-language corpora representing seven different countries. *Literary and Linguistic Computing* 22(1): 85–99. doi:10.1093/llc/fql044.

OBC = *Old Bailey Corpus*, version 0.4. Based on Tim Hitchcock, Robert Shoemaker, Clive Emsley, Sharon Howard and Jamie McLaughlin et al., *The Old Bailey Proceedings Online, 1674–1913*. Compiled by Magnus Huber and team at the Department of English, University of Giessen. http://www.uni-giessen.de/oldbaileycorpus/ (accessed 9 May 2014).

OED = *Oxford English Dictionary*. 1989. 2nd edition. OED Online. Oxford University Press. http://www.oed.com (accessed 9 May 2014).

Office for National Statistics 2004. *Focus on Social Inequalities*. http://www.ons.gov.uk/ons/rel/social-inequalities/focus-on-social-inequalities/2004-edition/ (accessed 9 May 2014).

Office for National Statistics 2013. *Women in the Labour Market*.

http://www.ons.gov.uk/ons/rel/lmac/women-in-the-labour-market/2013/ (accessed 9 May 2014).

Ogura, Mieko and William S.-Y. Wang 1996. Snowball effect in lexical diffusion: the development of *-s* in the third person singular present indicative in English. Derek Britton (ed.), *English Historical Linguistics 1994: Papers from the 8th International Conference on English Historical Linguistics, Edinburgh, 19–23 September 1994*, 119–141. Current Issues in Linguistic Theory 135. Amsterdam: John Benjamins.

Palander-Collin, Minna 1999. *Grammaticalization and Social Embedding: I THINK and METHINKS in Middle and Early Modern English*. Mémoires de la Société Néophilologique de Helsinki 55. Helsinki: Société Néophilologique.

Palander-Collin, Minna 2009. Variation and change in patterns of self-reference in early English correspondence. *Journal of Historical Pragmatics* 10(2): 260–285. doi:10.1075/jhp.10.2.06pal.

Palander-Collin, Minna and Mikko Hakala 2011. Standardized versions of the Corpora of Early English Correspondence. *Corpus Resource Database (CoRD)*. http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/standardized.html (accessed 9 May 2014).

Palmer, Chris C. 2009. Borrowings, derivational morphology, and perceived productivity in English, 1300–1600. PhD dissertation, University of Michigan. http://hdl.handle.net/2027.42/64624 (accessed 9 May 2014).

Paquot, Magali and Yves Bestgen 2009. Distinctive words in academic writing: a comparison of three statistical tests for keyword extraction. Andreas H. Jucker, Daniel Schreier and Marianne Hundt (eds.), *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*, 247–269. Language and Computers: Studies in Practical Linguistics 68. Amsterdam: Rodopi.

Parker, Kenneth 2004. Osborne, Dorothy [*married name* Dorothy Temple, Lady Temple] (1627–1695). Lawrence Goldman (ed.), *Oxford Dictionary of National Biography*. Online edition. Oxford: Oxford University Press. doi:10.1093/ref:odnb/27109.

PCEEC = *Parsed Corpus of Early English Correspondence*, tagged version. 2006. Annotated by Arja Nurmi, Ann Taylor, Anthony Warner, Susan Pintzuk, and Terttu Nevalainen. Compiled by the CEEC Project Team. York: University of York and Helsinki: University of Helsinki. Distributed through the Oxford Text Archive. http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/ (accessed 9 May 2014).

Pearson, Karl 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5* 50(302): 157–175. doi:10.1080/14786440009463897.

Pedersen, Ted 1996. Fishing for exactness. Computation and Language. Tom Winn (ed.), *Proceedings of the South-Central SAS Users Group Conference (SCSUG-96), Austin, TX, Oct 27–29, 1996*, 188–200. http://xxx.lanl.gov/abs/cmp-lg/9608010 (accessed 9 May 2014).

Pike, William A., John T. Stasko, Remco Chang and Theresa A. O'Connell 2009. The science of interaction. *Information Visualization* 8(4): 263–274. doi:10.1057/ivs.2009.22.

Plag, Ingo 1999. *Morphological Productivity: Structural Constraints in English Derivation*. Topics in English Linguistics 28. Berlin: Mouton de Gruyter.

Plag, Ingo 2003. *Word-Formation in English*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511841323.

Plag, Ingo 2006. Productivity. Bas Aarts and April McMahon (eds.), *The Handbook of English Linguistics*, 537–556. Blackwell Handbooks in Linguistics. Oxford: Blackwell.

Plag, Ingo, Christiane Dalton-Puffer and R. H. Baayen 1999. Morphological productivity across speech and writing. *English Language and Linguistics* 3(2): 209–228.

Plant, David 2005. Biography of John Jones. *British Civil Wars and Commonwealth Website*. http://www.british-civil-wars.co.uk/biog/john-jones.htm (accessed 9 May 2014).

Playfair, William 1801. *The Statistical Breviary*. London: T. Bensley.

Playfair, William 2005 [1786]. *Commercial and Political Atlas: Representing, by Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure, and Debts of England, during the Whole of the Eighteenth Century*. London: Corry. Republished in *The Commercial and Political Atlas and Statistical Breviary*, ed. by Howard Wainer and Ian Spence. Cambridge: Cambridge University Press.

Pohl, Nicole and Betty A. Schellenberg (eds.) 2003. *Reconsidering the Bluestockings*. San Marino, California: Huntington Library.

Pumfrey, Stephen, Paul Rayson and John Mariani 2012. Experiments in 17th century English: manual versus automatic conceptual history. *Literary and Linguistic Computing* 27(4): 395–408. doi:10.1093/llc/fqs017.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik 1989 [1985]. *A Comprehensive Grammar of the English Language*. London: Longman.

R Core Team 2014. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.r-project.org (accessed 9 May 2014).

Rainer, Franz 1988. Towards a theory of blocking: the case of Italian and German quality nouns. Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1988*, 155–185. Dordrecht: Foris.

Raumolin-Brunberg, Helena 1998. Social factors and pronominal change in the seventeenth century: the Civil-War effect? Jacek Fisiak and Marcin Krygier (eds.), *Advances in English Historical Linguistics (1996)*, 361–388. Trends in Linguistics: Studies and Monographs 112. Berlin: Mouton de Gruyter. doi:10.1515/9783110804072.361.

Raumolin-Brunberg, Helena and Terttu Nevalainen 2007. Historical sociolinguistics: the Corpus of Early English Correspondence. Joan C. Beal, Karen P. Corrigan and Hermann L. Moisl (eds.), *Creating and Digitizing Language Corpora, Volume 2: Diachronic Databases*, 148–171. Houndsmills: Palgrave Macmillan.

Rayson, Paul 2008. From key words to key semantic domains. *International Journal of Corpus Linguistics* 13(4): 519–549. doi:10.1075/ijcl.13.4.06ray.

Rayson, Paul, Dawn Archer, Alistair Baron, Jonathan Culpeper and Nicholas Smith 2007. Tagging the Bard: evaluating the accuracy of a modern POS tagger on Early Modern English corpora. Matthew Davies, Paul Rayson, Susan Hunston and Pernilla Danielsson (eds.), *Proceedings of Corpus Linguistics 2007, 27–30 July, University of Birmingham, UK*, article #192. http://www.corpus.bham.ac.uk/conference/proceedings.shtml (accessed 9 May 2014).

Rayson, Paul, Dawn Archer, Alistair Baron and Nicholas Smith 2008. Travelling through time with corpus annotation software. Barbara Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools, and Applications – State of the Art: PALC 2007*, 29–46. Łódź Studies in Language 17. Frankfurt am Main: Peter Lang. http://comp.eprints.lancs.ac.uk/id/eprint/2311 (accessed 9 May 2014).

Rayson, Paul, Damon Berridge and Brian Francis 2004. Extending the Cochran rule for the comparison of word frequencies between corpora. Gérald Purnelle, Cédrick Fairon and Anne Dister (eds.), *Le poids des mots (JADT 2004, Vol. 2) : actes des 7es journées internationales d'analyse statistique des données textuelles*, 926–936. Louvain-la-Neuve: Presses universitaires de Louvain. http://lexicometrica.univ-paris3.fr/jadt/jadt2004/tocJADT2004.htm (accessed 9 May 2014).

Rayson, Paul and Roger Garside 2000. Comparing corpora using frequency profiling. Adam Kilgarriff and Tony Berber Sardinha (eds.), *Proceedings of the Workshop on Comparing Corpora (WCC '00)*, 1–6. Stroudsburg, Pennsylvania: Association for Computational Linguistics. doi:10.3115/1117729.1117730.

Rayson, Paul, Geoffrey Leech and Mary Hodges 1997. Social differentiation in the use of English vocabulary: some analyses of the conversational component of the British National Corpus. *International Journal of Corpus Linguistics* 2(1): 133–152. doi:10.1075/ijcl.2.1.07ray.

Rayson, Paul, Andrew Wilson and Geoffrey Leech 2002. Grammatical word class variation within the British National Corpus Sampler. Pam Peters, Peter Collins and Adam Smith (eds.), *New Frontiers of Corpus Research: Papers from the Twenty First International Conference on English Language Research on Computerized Corpora, Sydney 2000*, 295–306. Language and Computers: Studies in Practical Linguistics 36. Amsterdam: Rodopi.

Renouf, Antoinette 2007. Tracing lexical productivity and creativity in the British media: 'the chavs and the chav-nots'. Judith Munat (ed.), *Lexical Creativity, Texts and Contexts*, 61–89. Studies in Functional and Structural Linguistics 58. Amsterdam: John Benjamins.

Renouf, Antoinette 2012. A finer definition of neology in English: the life-cycle of a word. Hilde Hasselgård, Jarle Ebeling and Signe Oksefjell Ebeling (eds.), *Corpus Perspectives on Patterns of Lexis*, 177–208. Studies in Corpus Linguistics 57. Amsterdam: John Benjamins.

Renouf, Antoinette and R. H. Baayen 1998. Aviating among the hapax legomena: morphological grammaticalisation in current British newspaper English. Antoinette Renouf (ed.), *Explorations in Corpus Linguistics: Proceedings of the 18th ICAME Conference, University of Liverpool, 21-25 May 1997*, 181–189. Language and Computers: Studies in Practical Linguistics 23. Amsterdam: Rodopi.

Riddle, Elizabeth M. 1985. A historical perspective on the productivity of the suffixes *-ness* and *-ity*. Jacek Fisiak (ed.), *Historical Semantics – Historical Word-Formation*, 435–461. Trends in Linguistics: Studies and Monographs 29. Berlin: Mouton de Gruyter.

Rissanen, Matti 2012. Grammaticalisation, contact and corpora: on the development of adverbial connectives in English. Irén Hegedűs and Alexandra Fodor (eds.), *English Historical Linguistics 2010: Selected Papers from the Sixteenth International Conference on English Historical Linguistics (ICEHL 16), Pécs, 23–27 August*

2010, 131–151. Current Issues in Linguistic Theory 325. Amsterdam: John Benjamins. doi:10.1075/cilt.325.

Roberts, Stephen K. 2004. Jones, John (*c*.1597–1660). Lawrence Goldman (ed.), *Oxford Dictionary of National Biography*. Online edition. Oxford: Oxford University Press. doi:10.1093/ref:odnb/15026.

Rohrdantz, Christian, Annette Hautli, Thomas Mayer, Miriam Butt, Daniel A. Keim and Frans Plank 2011. Towards tracking semantic change by visual analytics. Yuji Matsumoto and Rada Mihalcea (eds.), *HLT '11: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers – Volume 2*, 305–310. Stroudsburg, Pennsylvania: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=2002800 (accessed 9 May 2014).

Rohrdantz, Christian, Andreas Niekler, Annette Hautli, Miriam Butt and Daniel A. Keim 2012. Lexical semantics and distribution of suffixes: a visual analysis. Miriam Butt, Sheelagh Carpendale, Gerald Penn, Jelena Prokić and Michael Cysouw (eds.), *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*, 7–15. Stroudsburg, Pennsylvania: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=2388657 (accessed 9 May 2014).

Romaine, Suzanne 1982. *Socio-Historical Linguistics: Its Status and Methodology*. Cambridge Studies in Linguistics 34. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511720130.

Romaine, Suzanne 1983. On the productivity of word formation rules and limits of variability in the lexicon. *Australian Journal of Linguistics* 3(2): 177–200. doi:10.1080/07268608308599308.

Romaine, Suzanne 1985. Variability in word formation patterns and productivity in the history of English. Jacek Fisiak (ed.), *Papers from the 6th International Conference on Historical Linguistics, Poznań, 22–26 August 1983*, 451–465. Current Issues in Linguistic Theory 34. Amsterdam: John Benjamins.

Romaine, Suzanne 1998. Introduction. Suzanne Romaine (ed.), *The Cambridge History of the English Language, IV: 1776–1997*, 1–56. Cambridge: Cambridge University Press. doi:10.1017/CHOL9780521264778.002.

Saeed, John I. 1997. *Semantics*. Introducing Linguistics 2. Oxford: Blackwell.

Säily, Tanja 2005. Use of the suffixes *-ity* and *-ness* in early English letters: was gender a factor? Seminar paper, Department of English, University of Helsinki.

Säily, Tanja 2008. Productivity of the suffixes *-ness* and *-ity* in 17th-century English letters: a sociolinguistic approach. MA thesis, Department of English, University of Helsinki. http://urn.fi/URN:NBN:fi-fe200810081995 (accessed 9 May 2014).

Säily, Tanja 2011. Variation in morphological productivity in the BNC: sociolinguistic and methodological considerations. *Corpus Linguistics and Linguistic Theory* 7(1): 119–141. doi:10.1515/cllt.2011.006.

Säily, Tanja 2013. Progress in POS tagging the CEECE. *From Correspondence to Corpora: A Seminar on Digital Processing of Historical Letter Compilations, Helsinki, Finland, November 2013*. Seminar presentation. http://blogs.helsinki.fi/brevkonst/2013/10/17/from-correspondence-to-corpora-seminar-15-nov-2013/ (accessed 9 May 2014).

Säily, Tanja in press. Sociolinguistic variation in morphological productivity in eighteenth-century English. *Corpus Linguistics and Linguistic Theory*.

Säily, Tanja, Terttu Nevalainen and Harri Siirtola 2011. Variation in noun and pronoun frequencies in a sociohistorical corpus of English. *Literary and Linguistic Computing* 26(2): 167–188. doi:10.1093/llc/fqr004.

Säily, Tanja and Jukka Suomela 2007. Incidence matrices for *-ness* and *-ity* (CEEC, 17th century). Supplementary material to Säily and Suomela (2009). http://users.ics.aalto.fi/suomela/ity-ness-data/ (accessed 9 May 2014).

Säily, Tanja and Jukka Suomela 2009. Comparing type counts: the case of women, men and *-ity* in early English letters. Antoinette Renouf and Andrew Kehoe (eds.), *Corpus Linguistics: Refinements and Reassessments*, 87–109. Language and Computers: Studies in Practical Linguistics 69. Amsterdam: Rodopi.

Santorini, Beatrice 1990. *Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd Revision)*. University of Pennsylvania Department of Computer and Information Science. http://repository.upenn.edu/cis_reports/570/ (accessed 9 May 2014).

Santorini, Beatrice 2010. Annotation manual for the Penn Historical Corpora and the PCEEC. Department of Linguistics, University of Pennsylvania. http://www.ling.upenn.edu/hist-corpora/annotation/ (accessed 9 May 2014).

SCEEC = *Standardised-spelling Corpora of Early English Correspondence*. 2012. Compiled by Terttu Nevalainen, Helena Raumolin-Brunberg, Samuli Kaislaniemi, Jukka Keränen, Mikko Laitinen, Minna Nevala, Arja Nurmi, Minna Palander-Collin, Tanja Säily and Anni Sairio. Standardised by Mikko Hakala, Minna Palander-Collin and Minna Nevala. http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/ (accessed 9 May 2014).

Schröder, Anne 2008. On the productivity of verbal prefixation in English. Habilitationsschrift, Martin-Luther-Universität Halle-Wittenberg.

Scott, Mike 1997. PC analysis of key words – and key key words. *System* 25(2): 233–245. doi:10.1016/S0346-251X(97)00011-0.

Scott, Mike 2012. WordSmith Tools. Computer program. http://www.lexically.net/wordsmith/ (accessed 9 May 2014).

Shaffer, Juliet Popper 1995. Multiple hypothesis testing. *Annual Review of Psychology* 46: 561–584. doi:10.1146/annurev.ps.46.020195.003021.

Siirtola, Harri, Terttu Nevalainen, Tanja Säily and Kari-Jouko Räihä 2011. Visualisation of text corpora: a case study of the PCEEC. Terttu Nevalainen and Susan M. Fitzmaurice (eds.), *How to Deal with Data: Problems and Approaches to the Investigation of the English Language over Time and Space*. Studies in Variation, Contacts and Change in English 7. Helsinki: VARIENG. http://www.helsinki.fi/varieng/series/volumes/07/siirtola_et_al/ (accessed 9 May 2014).

Siirtola, Harri, Tanja Säily, Terttu Nevalainen and Kari-Jouko Räihä 2014. Text Variation Explorer: towards interactive visualization tools for corpus linguistics. *International Journal of Corpus Linguistics* 19(3): 417–429. doi:10.1075/ijcl.19.3.05sii.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability 26. Boca Raton, Florida: CRC / Chapman & Hall.

Sinclair, John 1991. *Corpus, Concordance, Collocation*. Describing English Language. Oxford: Oxford University Press.

Smitterberg, Erik 2008. The progressive and phrasal verbs: evidence of colloquialization in nineteenth-century English? Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta and Minna Korhonen (eds.), *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*, 269–289. Studies in Language Variation 2. Amsterdam: John Benjamins. doi:10.1075/silv.2.21smi.

Spence, Robert 2007. *Information Visualization: Design for Interaction*. 2nd edition. Harlow: Pearson Education.

Standop, Ewald 2000. Englische Verbkomplementation. *Anglia* 118(2): 217–257.

Štekauer, Pavol, Don Chapman, Slávka Tomaščíková and Štefan Franko 2005. Word-formation as creativity within productivity constraints: sociolinguistic evidence. *Onomasiology Online* 6: 1–55. http://www1.ku-eichstaett.de/SLF/EngluVglSW/OnOn6.htm (accessed 9 May 2014).

Suomela, Jukka 2007. *types*1: type and hapax accumulation curves. Computer program. ZENODO. doi:10.5281/zenodo.9860.

Suomela, Jukka 2014. *types*2: type and hapax accumulation curves. Computer program. ZENODO. http://users.ics.aalto.fi/suomela/types2/ (accessed 9 May 2014). doi:10.5281/zenodo.9868.

Szmrecsanyi, Benedikt forthcoming. About text frequencies in historical linguistics: disentangling environmental and grammatical change. *Corpus Linguistics and Linguistic Theory*.

Tagliamonte, Sali A. and Alexandra D'Arcy 2007. Frequency and variation in the community grammar: tracking a new change through the generations. *Language Variation and Change* 19(2): 199–217. doi:10.1017/S095439450707007X.

Tagliamonte, Sali A. and Alexandra D'Arcy 2009. Peaks beyond phonology: adolescence, incrementation, and language change. *Language* 85(1): 58–108. doi:10.1353/lan.0.0084.

Talbot, Mary 2006. Gender and language. Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 740–742. 2nd edition. Oxford: Elsevier. doi:10.1016/B0-08-044854-2/00331-X.

Taylor, Ann 2003. YCOE Lite: a beginner's guide to the York Corpus of Old English. University of York. http://www-users.york.ac.uk/~lang22/YCOE/doc/annotation/YcoeLiteToc.htm (accessed 9 May 2014).

Taylor, Ann 2007. The York–Toronto–Helsinki Parsed Corpus of Old English Prose. Joan C. Beal, Karen P. Corrigan and Hermann L. Moisl (eds.), *Creating and Digitizing Language Corpora, Volume 2: Diachronic Databases*, 196–227. Houndsmills: Palgrave Macmillan.

Taylor, Ann and Beatrice Santorini 2006. The Parsed Corpus of Early English Correspondence. University of York. http://www-users.york.ac.uk/~lang22/PCEEC-manual/ (accessed 9 May 2014).

Theus, Martin 2011. Mondrian – interactive statistical data visualization in Java. Computer program. http://stats.math.uni-augsburg.de/mondrian/ (accessed 9 May 2014).

Theus, Martin and Simon Urbanek 2008. *Interactive Graphics for Data Analysis: Principles and Examples*. Computer Science and Data Analysis. Boca Raton, Florida: CRC / Chapman & Hall.

Tieken-Boon van Ostade, Ingrid 2010. Eighteenth-century women and their norms of correctness. Raymond Hickey (ed.), *Eighteenth-Century English: Ideology and*

Change, 59–72. Studies in English Language. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511781643.005.

Tukey, John W. 1977. *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley.

Tweedie, Fiona J. and R. H. Baayen 1998. How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities* 32(5): 323–352. doi:10.1023/A:1001749303137.

Tyrkkö, Jukka 2013. "My intent Is onelie to further those that be willing to learne": the lexicon of mid-sixteenth-century surgical books in context. R. W. McConchie, Teo Juvonen, Mark Kaunisto, Minna Nevala and Jukka Tyrkkö (eds.), *Selected Proceedings of the 2012 Symposium on New Approaches in English Historical Lexis (HEL-LEX 3)*, 177–188. Somerville, Massachusetts: Cascadilla Proceedings Project. http://www.lingref.com/cpp/hel-lex/2012/abstract2846.html (accessed 9 May 2014).

Tyrkkö, Jukka 2014. "Strong churlish purging Pills": multi-adjectival premodification in early modern medical writing in English. Irma Taavitsainen, Andreas H. Jucker and Jukka Tuominen (eds.), *Diachronic Corpus Pragmatics*, 157–188. Pragmatics & beyond New Series 243. Amsterdam: John Benjamins. doi:10.1075/pbns.243.11tyr.

Urbanek, Simon, Tobias Wichtrey, Alex Gouberman and Martin Theus 2013. iPlots – interactive graphics for R. Computer program. http://www.iplots.org (accessed 9 May 2014).

Vartiainen, Turo, Tanja Säily and Mikko Hakala 2013. Variation in pronoun frequencies in early English letters: gender-based or relationship-based? Jukka Tyrkkö, Olga Timofeeva and Maria Salenius (eds.), *Ex Philologia Lux: Essays in Honour of Leena Kahlas-Tarkka*, 233–255. Mémoires de la Société Néophilologique de Helsinki 90. Helsinki: Société Néophilologique.

Ware, Colin 2004. *Information Visualization: Perception for Design*. 2nd edition. San Francisco: Morgan Kaufmann.

Warner, Anthony 2005. Why DO dove: evidence for register variation in Early Modern English negatives. *Language Variation and Change* 17(03): 257–280. doi:10.1017/S0954394505050106.

Welch, B. L. 1947. The generalization of "Student's" problem when several different population variances are involved. *Biometrika* 34(1–2): 28–35. doi:10.1093/biomet/34.1-2.28.

Wermser, Richard 1976. *Statistische Studien zur Entwicklung des englischen Wortschatzes*. Schweizer Anglistische Arbeiten 91. Bern: Francke.

Wilcoxon, Frank 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6): 80–83. http://www.jstor.org/stable/3001968 (accessed 9 May 2014).

Wolfson, Nessa 1990. The bulge: a theory of speech behavior and social distance. *Penn Working Papers in Educational Linguistics* 2(1): 55–83.

Wrightson, Keith 1993. *English Society, 1580–1680*. London: Routledge.

Appendix I: Glossary of statistical terms

absolute frequency The number of observations of an item in a corpus. Cf. *relative frequency*.

accumulation curve A graph showing how the number of observations grows as corpus size increases.

activation level A measure of *morphological productivity:* the number of *tokens* representing those *types* of a given affix whose frequency of occurrence is smaller than a given threshold. The level indicates how quickly the affix will be recognised and combined with the base (Baayen 1993: 195–196; Section 2.3.2).

bag-of-words model A representation where all words are assumed to be statistically independent (Section 7.4.1).

bag-of-words test *Significance tests* of this kind make the assumption that words occur randomly in texts (Section 5.1.1). Examples include the χ^2 test, the log-likelihood ratio test, Fisher's exact test, the binomial test and the test of equal or given proportions. Cf. *dispersion-aware test*.

Bernoulli trial A random event with a binary (yes/no) outcome (Section 7.4.1).

bimodal distribution A distribution with two peaks.

binomial distribution The probability of getting a certain number of 'yes' outcomes in a sequence of *Bernoulli trials*.

binomial test A *bag-of-words test* (when applied to comparing word frequencies).

Bonferroni correction A method for estimating the *family-wise error rate* when testing multiple hypotheses (Section 5.1.3).

bootstrap test A *dispersion-aware test* based on *bootstrapping* (Section 7.4.2).

bootstrapping Drawing samples randomly with replacement, so that an individual sample may be drawn more than once; see *resampling*.

burstiness The phenomenon that the occurrences of a word tend to cluster together in a corpus (Section 7.2.1).

category-conditioned degree of productivity See *potential productivity*.

chi-square test See *χ^2 test*.

collocation Recurrent and predictable co-occurrence of words (Evert 2009: 1214).

confidence interval An estimate of the uncertainty in an observed frequency (Section 5.1.2).

confirmatory analysis Analysis performed in order to test pre-specified hypotheses.

contingency table A table showing the frequency distribution of variables (e.g., the frequencies of linguistic features in different corpora).

dis legomenon A word occurring twice in a given data set (Section 2.3.2).

dispersion How evenly a linguistic feature is distributed across the texts in a corpus (Section 5.1.1).

dispersion-aware test A *significance test* that takes into account the *dispersion* of the linguistic feature in the corpus (Section 5.1.1). Cf. *bag-of-words test*.

distribution See *probability distribution*.

effect size A measure of how large a difference is, irrespective of *statistical significance* (Section 5.1.1).

expanding productivity The ratio between the number of *hapaxes* with a given affix (n_1) and the total number of all hapaxes (h) in the corpus: $P^* = n_1/h$. This measure indicates how much the affix contributes to the overall vocabulary growth of the corpus (Baayen 1993: 193; Section 2.3.2).

extent of use See *realised productivity*.

false discovery rate control The expected proportion of *false positives* out of all positives when testing multiple hypotheses (Section 5.1.3).

false negative A result erroneously marked as non-significant.

false positive A spurious result marked as significant (Section 5.1.1).

family-wise error rate The probability that at least one *false positive* occurs when testing multiple hypotheses (Section 5.1.3).

Fisher's exact test A *bag-of-words test* (when applied to comparing word frequencies).

frequency See *absolute frequency*, *relative frequency*.

G^2 test See *log-likelihood ratio test*.

growth curve See *accumulation curve*.

h The total number of *hapaxes* in the corpus.

hapax See *hapax legomenon*.

hapax legomenon A word occurring only once in a given data set (Section 2.3.2).

hapax-conditioned degree of productivity See *expanding productivity*.

hypothesis A testable assumption made by the researcher (Section 5.1).

incidence matrix A table showing which items (e.g., words of interest) occur in each text (Section 6.3.1).

key word analysis A procedure for establishing which words occur at an unusual frequency in a (sub)corpus compared with a reference corpus (Scott 1997).

linear Directly proportional.

log-likelihood ratio test A *bag-of-words test* (when applied to comparing word frequencies); see further Section 7.4.1.

Mann-Whitney ranks test See *Wilcoxon rank-sum test*.

Mann-Whitney U test See *Wilcoxon rank-sum test*.

matrix A two-dimensional table of values. Cf. *vector*.

mean Average; the sum of elements divided by the number of elements.

median The middle value in an ordered list of values: half of the values are above and half below the median.

Monte Carlo sampling Picking a number of objects at random from a suitable *probability distribution*.

Monte Carlo testing A method by which one picks elements using *Monte Carlo sampling*, checks which percentage of them satisfies the desired properties, and derives an estimate of the total number of such objects (Section 6.5.2).

morphological productivity "The statistically determinable readiness with which an element enters into new combinations" (Bolinger 1948: 18); a multifaceted phenomenon that can be formalised in terms of *realised*, *potential* and *expanding productivity* (Baayen 2009; Section 2.3.2).

multimodal distribution A distribution with multiple peaks (Section 5.2.3).

N The number of *tokens* with a given affix in the corpus.

n_1 The number of *hapaxes* with a given affix in the corpus.

nonparametric Statistical methods that do not use *parametric models*.

normal distribution Gaussian *probability distribution* (bell-shaped curve).

normalised frequency See *relative frequency*.

null hypothesis The assumption that an observation is explainable through chance. Assumed to be true, tested (see *significance test*) and rejected if there is enough evidence (Section 5.1).

one-sided test A test that is only concerned with deviation in one, pre-specified direction. As an example, the hypothesis could be that the frequency of an item is exceptionally low (Section 6.4.2). Cf. *two-sided test*.

outlier An atypical data point (e.g., distant from others).

P See *potential productivity*.

P* See *expanding productivity*.

***p*-value** The probability that we are wrong in rejecting the *null hypothesis*.

parametric model A model assuming that the phenomenon under analysis follows a *probability distribution* involving a finite number of parameters.

permutation Reordering.

permutation testing Drawing samples randomly without replacement, so that no individual sample occurs more than once; see *resampling*. Also used as a *dispersion-aware test* (Sections 6.4.2–6.4.3).

post-hoc correction A correction applied to *p-values* after testing multiple hypotheses. Examples include *Bonferroni correction* and *false discovery rate* control.

potential productivity The ratio between the number of *hapaxes* with a given affix (n_1) and the total number of *tokens* with that affix (N) in the corpus: $P = n_1/N$. This measure expresses the probability of observing new types with the relevant affix when N tokens with the affix have been sampled (Baayen and Lieber 1991: 809–810; Section 2.3.2).

probability distribution Specifies a probability for each possible observation. Examples: *uniform distribution*, *normal distribution*, *binomial distribution*.

quartiles Points below which there are 1/4, 2/4 and 3/4 of the data values.

quintiles Points below which there are 1/5, 2/5, 3/5 and 4/5 of the data values.

realised productivity The number of different words (i.e., *types*) V with a given affix in the corpus (Baayen 2009; Section 2.3.2).

relative frequency The number of observations of an item normalised to (or divided by), e.g., the total number of words in the corpus. Cf. *absolute frequency*.

resampling Drawing samples from a corpus repeatedly and randomly to calculate *confidence intervals* for the observed frequency of an item in a subcorpus (Section 5.1.2). Examples of resampling statistics include *bootstrapping* and *permutation testing*.

robust statistics Statistics requiring a minimal number of background assumptions (Section 5.1.1).

significance level See *statistical significance*.

significance test A procedure for calculating the probability that we are wrong in rejecting the *null hypothesis* (Section 5.1).

significance threshold See *statistical significance*.

statistical significance An observed difference is statistically significant if the probability that we are wrong in rejecting the *null hypothesis* is lower than a specific percentage, called the significance level or the significance threshold (Section 5.1).

t-test A *dispersion-aware test* entailing the assumption that the sample means follow *normal distributions* (Section 13.2).

test of equal or given proportions A *bag-of-words test* (when applied to comparing word frequencies).

token An occurrence of a word in a text. In a text of 1,000 running words, there are 1,000 word tokens (Section 2.3.2). Cf. *type*.

Tukey's test A *dispersion-aware test* that is similar to the *t-test* but incorporates a measure of *family-wise error rate* (Section 5.1.3).

two-sided test A test that is concerned with deviation in either direction. Cf. *one-sided test*.

type A word in the abstract sense; an individual word-form or lexeme (Section 2.1). If there are 300 different words in a text of 1,000 running words, the number of word types is 300 (Section 2.3.2). Cf. *token*.

type accumulation curve A graph showing how the number of *types* grows as corpus size increases (Section 5.2.2).

type I error See *false negative*.

type II error See *false positive*.

uniform distribution (discrete) A *probability distribution* in which each observation is equally probable (Section 6.5.2).

V See *realised productivity*.

vector A one-dimensional table (i.e., a list) of values.

Wilcoxon rank-sum test A *dispersion-aware test* based on the ranks of the samples ordered by frequency (Section 13.2).

χ^2 test A *bag-of-words test* (when applied to comparing word frequencies).

Appendix II: Chief sociolinguistic parameters of the *Corpora of Early English Correspondence*

The present work focuses on the first three categories listed below, i.e., gender, social rank and the relationship between the sender and the recipient of the letter. The parameters also include date of birth and the date of the letter, which makes it possible to calculate the age of the correspondents at the desired level of granularity. In addition, each correspondent has a unique ID to facilitate the study of individuals. For more information, see the entry for CEEC in the Corpus Resource Database (CoRD).

Gender

F	Female
M	Male

Social rank

R	Royalty
N	Nobility
GU	Upper gentry
GL	Lower gentry
CU	Upper clergy
CL	Lower clergy
P	Professionals
M	Merchants
O	Other non-gentry

Relationship between sender and recipient

FN	Nuclear family
FO	Other family
FS	Family servants
TC	Close friends
T	Other acquaintances

Social mobility

U	Up
D	Down
N	None

Education

H	Higher
C	Cambridge

O	Oxford
I	Inns of Court
F	Foreign
A	Apprentice
E	Elementary
PC	Private, classical
PN	Private, non-classical

Place of birth; main domicile

N	North
F	East Anglia
H	Home Counties
L	London
C	Court
A	Abroad

Migration history

Y	Yes
YL	Yes, London
YA	Yes, abroad
YLA	Yes, London and abroad