The Emergence of Group Stress in Medieval French

Thomas Michael Rainsford
St. John's College

24 January, 2011

This dissertation is submitted for the degree of
Doctor of Philosophy.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

Statement of length

The word count of this dissertation does not exceed 80,000 words including footnotes, references and appendices, but excluding the bibliography.

Summary

The thesis investigates the development of the group-stress system in French from the earliest textual records to 1500. Empirical work is based on a corpus compiled especially for the study, which comprises 87 extracts from medieval French texts totalling over 250,000 words, composed mainly of verse texts to make use of the extra phonological information provided by the form. A unique metrical and syntactic annotation is used in the corpus to permit studies of phonological phrasing and stress placement in lines of verse.

Much octosyllabic narrative verse, in particular texts associated with oral performance, is shown to have an iambic rhythmic tendency in the pre-1250 period, which is particularly strong in the earliest texts. No such effect is found in lyric texts or plays, or in narrative from after 1250. Additionally, a phonological phrase boundary is commonly found in the middle of the line. Iambic rhythmic organization is argued to be incompatible with group stress and associated 'stress deafness' effects observed in modern French. From this data, group stress is argued to have developed between the mid-12$^{\text{th}}$ and mid-13$^{\text{th}}$ centuries.

Work on modern French (e.g. Post, 2000) suggests that the stress group is the phonological phrase. Through reconstruction of the phonological phrasing of medieval French, the thesis demonstrates that regular word-final stress, the phonological phrase internal process of stress clash resolution, and the frequency of monosyllabic words combine to favour reanalysis of the French stress system in the pre-1250 period. Finally, the hypothesis that prosodic change affected verb-second word order in medieval French is reconsidered. It is argued that light clause-initial constituents which do not form their own phonological phrase (i.e. short adverbs, subject pronouns) become unstressed, a development which triggers syntactic changes that lead to the introduction of non verb-second word orders.

Contents

Acknowledgements

This thesis owes a tremendous amount to the insight and encouragement of Professor Wendy Bennett. As a supervisor, she has never ceased to encourage me to tackle the most difficult problems and questions, reading the often rebarbative results of my efforts with astonishing speed, and providing feedback which pinpoints the many problems with remarkable clarity. I am deeply grateful for her support.

The research for this thesis was funded by the Arts and Humanities Research Council, grant 2007/133103. The metrical and phonological aspects have benefited from the comments of Francis Nolan, Bert Vaux and Nigel Fabb; while the discussion of historical syntactic change owes much to discussions with Ian Roberts, David Willis, Ioanna Sitaridou and Adam Ledgeway. The design of the corpus annotation benefited from discussions with the members of the *Syntactic Reference Corpus of Medieval French* project, in particular Sophie Prévost, Céline Guillot, Serge Heiden, Beatrice Bischof, Alexey Lavrentiev and Nicolas Mazziotta. I would also like acknowledge the assistance of Anne Cobby in helping me to source some of the corpus texts, and Hayley Fisher for her help with the statistical analysis. Any errors are, of course, my own.

My thanks go also to my graduate peers in the Linguistics department, Alastair Appleton, Ruth Cumming, Hae-sung Jeon, Petros Karatsareas, Elliott Lash, Erica Ross, Anna Tristram, George Walkden and Liv Walsh for friendly feedback on the ideas presented in this thesis.

On a personal note, I would like to offer special thanks to Hayley Fisher, Alexey Lavrentiev, Andrew Macdonald and Anna Tristram for putting me up (and putting up with me) in the difficult final stages. Finally, this thesis could never have been completed without the tremendous love and support of my family, in particular my sister Clare and my parents Michael and Helen Rainsford. I cannot even begin to express my gratitude.

Key to abbreviations

BFM	*Base de français médiéval*
DEAF	*Dictionnaire étymologique de l'ancien français*
MCVF	*Modéliser le changement: Les Voies du français*
MedFr	Medieval French
ModFr	Modern French
NCA	*Nouveau Corpus d'Amsterdam*
PhPh	Phonological phrase
PWd	Prosodic word
WMM	Weighted match mean

Key to metrical and phonological symbols

.	Syllable boundary (within a word)
//	Cæsura
a͡bc	Elision
(abc)	Uncounted syllable
abc	Stress
-	No phonological phrase boundary (between words)
/	Phonological phrase boundary

Introduction

The earliest description of the accentual system of French is to be found in John Palsgrave's *Esclaircissement de la langue francoyse*, written for English learners of French in 1530: [1]

> **For the kepyng of trewe accent**
> And noew to speke of theyr thyrde poynt / where I have shewed that the frenchemen studye to gyue euery worde that they abide and reste vpon / theyr most audible sou[n]de. The hole reason of theyr acce[n]t is gro[n]ded chefely vpon thre poyntes / fyrst there is no worde of one syllable whiche with them hath any accent / or that they vse to pause vpon / and there is one great cause why theyr tong semeth to vs so brefe and sodayn and so harde to be understa[n]ded whan it is spoken / especially of theyr paysantes or co[n]men people / for thoughe there come neuer so many wordes of one syllable together / they pronounce them nat distinctly asonder as the latines do / but sounde them all vnder one voyce and tenour / and neuer rest nor pause vpon any of them / except the co[n]myng next vnto a poynt be the cause therof. Seconde / euery worde of many syllables hath his accent upon the last syllable / but yet that nat withstandynge they vse vpon no suche worde to pause / except the co[n]myng next vnto a poynt be the causer therof / and this is one great thyng whiche inclineth the frenchemen so moche to pronounce the latin tong amysse / whiche co[n]trary neuer gyue theyr accent on the last syllable. The thyrde poynte is but an exception from the seconde / for whan the last syllable of a frenche worde endeth in E / the syllable next afore him must haue the accent / and yet is nat this rule euer generall / for if a frenche worde ende in Te / or haue z/ after E / or be a preterit partyciple of the fyrst coniugation / he shall haue his accent vpon [the] last syllable / accordyng to the seconde rule /
>
> (Palsgrave, 1530 [2003]: 56–57)

1. On the lack of such documents from the medieval period, see Marchello-Nizia (1995: 175).

The majority of Palsgrave's observations would be familiar to English learners of French today. The fundamental characteristic of French stress, he notes, is that the most prominent stress is realized on words which speakers 'rest upon'. This 'pausing' upon a word, Palsgrave continues, is associated with the presence of a punctuation mark: speakers do not pause on words unless 'the co[n]myng next vnto a poynt be the cause therof'. Writing for English-speaking students of French over 450 years later, Tranel makes a similar observation about modern French stress:

> One can say that within each syntactically (and semantically) delimited portion of a sentence, it is normally the word furthest to the right which stands out.
>
> (Tranel, 1987: 199)

Tranel refers to these syntactic delimiters as 'syntactic breaks' (1987: 198). It seems valid to suggest that what Palsgrave describes as speakers 'pausing upon' a word before a punctuation mark and giving it 'theyr most audible sou[n]de', Tranel describes as a syntactically delimited 'break' in the discourse, directly preceded by a word which 'stands out'. All words of more than one syllable bear final stress. Most interestingly, Palsgrave observes that French speakers extend the final-stress pronunciation of their own language to Latin. This oxytonic pronunciation may suggest that the fixed final stress of French made it more difficult for speakers to realize stress in any other position, even in a language with a more variable stress rule. Lastly, Palsgrave notes that most words spelt with a final <e> bore penultimate stress.[2] In this respect, 16th-century French differs from the standard French of today, from which all post-tonic syllables have disappeared.[3] Thus, 'stress always falls on the last pronounced syllable' (Tranel, 1987: 33). Tranel's reference to the 'last pronounced syllable' reminds us that former post-tonic syllables are still shown in modern French orthography.

The parallels between the accounts of Palsgrave and Tranel suggest that the stress system of the early 16th century was not greatly different from that of modern French. The system is a 'group stress' system, where primary stress falls on the final (non-schwa) syllable of a group of words, and words within the stress group have no primary stress.

2. Except for the <-ez> ending, all stressed final <e>s described by Palsgrave would be marked with an acute accent in modern French orthography.

3. However, pronunciation of final schwa remains common in southern French (cf. Walter, 1982: 93–95).

It is not clear when this group-stress system first appeared in French. However, historians of French assume that a very different system was present in the Gallo-Roman period. A long-standing hypothesis is that contact with Frankish, following the settlement of the Franks in northern Gaul from the late 5th century onwards, led to a *strengthening* of the stress on individual words in Gallo-Romance. Pope describes the development as follows:

> The Germanic accent was a strong expiratory one, and its influence intensified considerably the stress accorded to the *tonic* syllable and lessened in proportion that placed on the others [...] thus the relatively level accentuation of Early Gallo-Roman was changed into a more up and down system, very much like that of Modern English.
>
> (Pope, 1952: §223)

The English parallel here is also reflected in the sound changes which Pope attributes to strong tonic stress: stressed mid-vowels in open syllables form diphthongs, while unstressed vowels are reduced to schwa or deleted, 'all phenomena that are closely paralleled in Modern English' (Pope, 1952: §223). If Pope is correct, the stress system which Palsgrave found so alien in the 16th century was in fact very English-like a thousand years earlier. There can be little doubt that a radical prosodic change occurred in the intervening period, fundamentally altering the nature of stress in the language.

1 Views of prosodic change

No major study exists dedicated solely to the history of stress in French, and there is a clear need for further research:

> De ce changement accentuel, de l'époque où il s'est produit, des étapes par lesquelles il est passé, de ses causes, on ne sait à peu près rien avec certitude.
>
> (Marchello-Nizia, 1995: 175)

However, a number of historians of French have placed emphasis on the importance of prosodic change in the development of the phonology, morphology and syntax of the language. Such a selection of 'les hypothèses les plus diverses' (Marchello-Nizia, 1995: 175) hardly constitutes a 'literature' on medieval French prosody. Nevertheless, in order to give an overview of the problem, I will compare four views of prosodic change from scholars of the history of French (Pope, 1952; Ewert, 1961; Kukenheim, 1971; Marchello-Nizia, 1995)

and one view from a scholar specializing in the development of the Romance languages (Banniard, 1998).

Throughout her study of the development of the historical phonology and morphology of French, Pope (1952) places great emphasis on the importance of stress:

> The dominant factors in the evolution of pronunciation in Later Old and Middle French [end of the 11th century–end of the 16th century] are the gradual lessening of the heavy tonic stress that characterised Period I [end of the 5th century–end of the 11th century] and a new tendency to link closely together words closely connected in thought.
>
> (Pope, 1952: §170)

For Pope, change in prosody is the 'dominant factor' in the evolution of pronunciation. It is *causative*: sound changes such as the monophthongization of diphthongs and the loss of consonantal enclitic pronouns 'ensue' (1952: §171) from prosodic change. Prosodic change itself is divided into two parts: a phonetic development (the weakening of stress) and a syntactic-phonetic development (the 'linking together' of words). Chronologically, these developments take place gradually from the end of the 11th century up until the end of the 16th century.

Ewert (1961: §§145–49), on the other hand, argues that prosodic change may be characterized as a shift from 'word stress' to 'group stress', although it is not clear how he understands these terms.[4] While for Pope prosodic change causes phonological change, for Ewert the final emergence of group stress is the consequence of a phonological change: the loss of final schwa in the 16th century. Ewert argues that the loss of unstressed post-tonic schwa led to an increase in the proportion of words consisting only of historically stressed syllables. Since stress cannot be realized on every syllable in a group of words, the stress on individual words was equalized. Chronologically, while Ewert finds some evidence for group stress in the 12th and 13th centuries, it is 'impossible to assign a definite date' to the emergence of the new accentual system other than that 'it was constituted by the end of the sixteenth century' (1961: §147).

Kukenheim (1971) adopts a third different view of prosodic change: it is a general shift from descending to ascending rhythm. Descending rhythm, inherited from Latin, is characterized in Kukenheim's view both by the presence

4. Word stress is glossed simply as 'the alternation of tonic and atonic syllables' (1961: §145), and there is no attempt to define the 'group' of 'group stress'.

of post-tonic syllables in Latin words (e.g. MA.TER, SO.LI.CU.LUM) [5] and by a tendency for the first word in a group of words to be emphasized (e.g. GRAECOS VICERUNT ROMANI 'the Romans even conquered the Greeks' vs. ROMANI VICERUNT GRAECOS 'it was the Romans who conquered the Greeks') (1971: 318). By contrast, from the late 13th century French had developed an ascending rhythm, characterized at the word level by a regular final stress (or penultimate before final schwa) and at the group level by a number of changes in word order. [6] There is no sense in Kukenheim's analysis that a 'group stress' replaces a 'word stress': the accentuation of both words and groups is argued to be governed by a single rhythmic development. Kukenheim argues that the period between the 7th century and the mid-13th century was one of rhythmic fluctuation, but that by the 14th century 'la langue française s'était complètement débarrassée de l'intonation descendante' (1971: 321). Like Pope, Kukenheim considers rhythmic change to be a causal factor in linguistic change:

> Prenant l'hypothèse de ce 'renversement du rythme' comme point de départ, nous allons tâcher de montrer quelles [en] ont été les conséquences [...] sur le comportement des sons, des formes et des syntagmes.
>
> (Kukenheim, 1971: 321)

The range of phonological, morphological and syntactic changes identified by Kukenheim is considerable, from the simplification of word-initial affricates to the relative ordering of nouns and adjectives.

Marchello-Nizia (1995), like Ewert, considers prosodic change to be characterized by a shift from word to group stress:

> En français [...] il y a eu, et c'est l'un des changements importants qu'a subis cette langue, le *passage d'une accentuation de mot à une accentuation de groupe de mots.* On est passé d'un état où chaque mot ou presque (le latin connaissait quelques clitiques) avait une autonomie accentuelle, où la frontière entre les mots était discernable, où surtout l'accent avait un rôle distinctif, à un état où dans l'énoncé les mots sont liés prosodiquement selon des règles 'de liaison' pour la plupart strictes et bien décrites pour le français contemporain.
>
> (Marchello-Nizia, 1995: 174; emphasis original)

Marchello-Nizia's position is also similar to that of Pope, as she highlights the emergence of the 'linking' of words as an inherent feature of prosodic change.

5. Underlining is used to represent stress.

6. For example, the generalization of unstressed object pronouns in the (now unstressed) pre-verbal position (Kukenheim, 1971: 320).

However, in Marchello-Nizia's view, the change is additionally connected to the loss of 'distinctive' stress: i.e. the emergence of a regular word-final stress rule. Like Kukenheim, Marchello-Nizia identifies a period of overlap between old and new stress systems. In the 12th and 13th centuries, she suggests that a regular group-final stress could be accompanied by an emphatic group-initial word stress. Oxytonic, word-final stress had clearly developed by this point. However, a number of morpho-syntactic phenomena (cf. chapter one, §3.2) seem to indicate that the *initial* position in the clause and perhaps also the noun phrase was potentially stressable and emphatic (Marchello-Nizia, 1995: 180). This idea of a strong group-initial position is similar to the 'descending rhythm' of Kukenheim's analysis.

Finally, Banniard (1998) suggests that there was a weakening of word stress. In essence, he suggests that stress in Latin plays a crucial role in facilitating the segmentation of the speech signal into words. A stressed syllable, he implies, serves to signal an upcoming word boundary:

> La prédictabilité d'une frontière de mot dépend du rapport entre l'accentuation effective du mot dans la réalisation langagière et sa place entre son prédécesseur et son successeur immédiats. La succession accents / frontières de mots se déroule dans l'ordre syntagmatique.
>
> (Banniard, 1998: 69)

However, with the loss of the majority of post-tonic syllables in the 8th century, stressed syllables and word boundaries frequently come to coincide. This simpler stress pattern, Banniard argues, requires less emphasis to be placed on the stressed syllables, as regular oxytonic forms have introduced 'une nouvelle sécurité discriminative dans l'ordre syntagmatique' (1998: 70). This argument appears to me rather counter-intuitive: indeed, it could equally well be argued that the demarcative function of stress is strengthened by its regular occurrence at word boundaries, rather than becoming 'redundant', as Banniard suggests. [7] Chronologically, Banniard argues for a weakening of tonic stress from the 9th century, with a subsequent development, perhaps linked to the loss of post-tonic schwa, leading to its eventual disappearance by the modern French period.

Faced with these diverse views, it is apparent that there is little consensus as to the nature of the prosodic change in French. For Banniard, it is simply a weakening of stress. For Pope, the weakening of stress is accompanied by the

7. On the demarcative function of modern French stress, see, for example, Lacheret-Dujour and Beaugendre (1999: 43).

linking together of words into groups. For Ewert and Marchello-Nizia, the primary characteristics of the change are the linking together of words into groups and the loss of the accentual independence of the word. Kukenheim, however, takes an entirely different view: the change is simply a 'rhythmic reversal', affecting words and groups of words equally. There is also little consensus about the dating of the change. Kukenheim and Banniard identify change as beginning with the emergence of regular word stress in the 7th century; Ewert instead chooses to focus on the loss of all post-tonic schwa vowels in the 16th century. Kukenheim and Marchello-Nizia consider the period until the 13th century as one of fluctuation, or of the co-existence of two systems. Pope argues that prosodic change progresses gradually from the end of the 11th century to the end of the 16th century. Discounting Ewert's estimate,[8] if consensus can be found here, it is that change is in progress over the 12th and 13th centuries. Finally, the studies show important differences in emphasis. Ewert and Banniard focus primarily on the causes of prosodic change, while Pope and Kukenheim focus on the effects that prosodic change had on the linguistic system. This seems to reflect a different view of the causality of linguistic change. Thus, for Pope, the weakening of heavy tonic stress seems to have no language-internal cause, but is responsible for a wide range of phonological changes (e.g. the levelling of diphthongs; 1952: §508). For Banniard and Ewert, on the other hand, it is segmental phonological changes which cause the weakening of stress. Yet there is no reason to adopt a fixed directionality of change here, and in the present thesis, I take the view that prosodic change has a language-internal cause, but may itself trigger wider changes in the linguistic system.

2 Research questions

Four core questions may be identified to which a successful study of prosodic change must provide answers:

Character: What is it that changes in the medieval French stress system?

Chronology: When does the change take place? If in several stages, what are the stages?

Mechanism: How does the change take place? What are the features of medieval French that favour a prosodic change?

8. Palsgrave's testimony suggests that group stress existed in the 16th century despite the presence of post-tonic schwa.

Consequence: What are the consequences (if any) of prosodic change for other aspects of the language?

The present thesis focuses on the emergence of group stress as the primary characteristic of prosodic change in the medieval period. There are a number of reasons for this. Firstly, the idea of 'group stress' is widely accepted in studies of modern French (e.g. Lacheret-Dujour and Beaugendre, 1999: 43). Group stress can be formalized in modern linguistic terms, unlike the 'weak tonic stress' of Pope or the 'ascending rhythm' of Kukenheim.[9] Secondly, it is the presence of a group-stress system which most clearly differentiates modern French from other Romance varieties, and thus by implication from an earlier stage in its own history. The association of stress with a 'pause' or 'break' at the end of a group of words is noted by Palsgrave in the 16th century and by Tranel in the 20th century; neither comment on whether stress is 'weak' or 'strong', as Pope's characterization might suggest. The reality of the word vs. group stress distinction is not in doubt, and nor is the fact that group stress in French develops from a word-stress system. However, while the general character of the change to be studied is clear, the properties of 'group stress' must be better understood and more clearly formalized. To this end, chapter one begins with a review of the properties of the modern French stress system, with regard both to the definition of the 'group' and its interaction with the position of stress.

The second focus of the study is the 'chronology' problem. There is some consensus among previous studies that the emergence of group stress can be dated to the medieval period, perhaps with a focus on the 12th and 13th centuries. We can certainly suppose a *terminus ad quem* for the emergence of group stress, and consequently for the present study, at the beginning of the 16th century, shortly before Palsgrave's treatise was written. Establishing a *terminus a quo* for the change is less straightforward. Kukenheim, Banniard and Marchello-Nizia suggest that the loss of word stress may begin following the emergence of a regular stress rule. In the second part of chapter one, I consider stress-conditioned changes at the word level in the history of French. These provide a *terminus a quo* for the emergence of group stress. Since the *terminus a quo* remains before the textual period on the basis of evidence from historical phonological changes, I argue that a different type of evidence is necessary if we are to fix the emergence of group stress more accurately within the

9. It is possible that 'heavy stress' could be interpreted as 'strongly stress-timed'; cf. chapter one, note 39. Jacobs (1993) suggests a formalization of 'ascending rhythm', but only with reference to a single aspect of medieval French morpho-phonology.

medieval period. In particular, I claim that the use of rhythmic organization in versification provides an insight into how speakers *perceive* stress in their language, and as such is one of the most useful sources of historical data that can be used to study prosodic change. Chapter three investigates changes in the rhythmic organization of verse in the medieval period, with the primary goal of pinpointing the date of prosodic change.

Because of the focus on accurate chronology in the present study, I have chosen to refer to the period from the earliest textual records in the 9th century to 1500 as 'medieval French' (MedFr). Datings within the period are given by reference to a specific century or centuries. Period labels, such as 'Old French' or 'Middle French' are avoided. Not only does this provide less scope for misinterpretation, but it emphasizes the linguistic continuity of French in the medieval period.

The two remaining problems, the mechanism and consequences of change, cannot be tackled until the character and chronology have been established. It is not possible to show that a particular set of linguistic conditions causes the emergence of group stress unless the date at which the change occurs is known. It is clear from the outset that Ewert's 16th-century mechanism of change is problematic, since it post-dates the *terminus ad quem* we have adopted based on Palsgrave's testimony. Equally, if prosodic change is argued to trigger further linguistic changes, these further changes must be attested after prosodic change has taken place, and there must be a plausible causal link between the two. These problems will form the subject of chapter four, in which I will argue that prosodic change had a phonological cause, and morpho-syntactic effects.

3 Use of written sources

Any study of a medieval language must be based on written textual records. The use of such evidence imposes obvious constraints: we cannot study the spoken form of the language, we have no access to native speakers' intuitions, we cannot fill any gaps in the textual record and we have less control over factors which may cause linguistic variation.

Few areas of historical study can be as seriously challenged by the nature of the data as the study of stress. In MedFr, stress is a non-phonemic feature of the spoken form of the language, and is not represented in any written text. Stress is thus not a property of the written form of the language on which

historical studies must be based.

One approach to this problem is to study stress through the study of stress-conditioned linguistic features which are recorded in the texts. In French, phonological changes such as diphthongization and vowel reduction are conditioned by the position of primary word stress. For this approach to be successful, there must be no doubt that the phenomenon is stress-conditioned. Changes in morpho-syntax, such as the form and position of pronouns and the verb-second word order of MedFr, have been used in previous studies as evidence for a change in the stress system. However, as we will see in chapter one (§3.2), the role of stress here is convincingly disputed.

In the present thesis, stress is studied more directly through partial reconstructions of the stress system of MedFr. This is possible since some aspects of the French stress system remain unchanged by the emergence of group stress. Primary stress in modern French is only realized on etymologically stressed syllables. [10] A possible reconstruction of the position of stress at any period in the history of French can therefore be based on the assumption of an underlying stress on the final, etymologically stressed, syllable. This type of reconstruction is used in chapter three in order to study the rhythm of verse texts. The domain of stress group is also argued to be constant. In chapter one (§1), I will argue for the view that the stress group in modern French is a phonological phrase, a prosodic constituent whose definition is constant across all languages (Nespor and Vogel, 1986: 168–69). Reconstruction of stress groups is used particularly in chapter four, where I will focus on the causes and the effects of the loss of primary stress from all non group-final words.

In reconstructing the stress pattern of a text, I implicitly assume that the written text has a spoken form. In the case of MedFr, this assumption would seem to be valid, since texts were generally read aloud (cf. Marnette, 1998: 169–70, note 10). Nevertheless, the link to the spoken form is not equally strong in all texts. Legal documents such as charters have weak links to orality, since their primary purpose is to serve as a written record of events, not to be read aloud. On the other hand, verse epics (*chansons de geste*) are associated with an oral storytelling tradition. The written form here is intended to be sung; indeed, the text may even have been composed orally and only subsequently written down (cf. chapter two, §1.3.1) Although not all verse texts are as strongly linked to oral performance as the *chansons de geste*, spoken features

10. With the important exception of post-verbal pronouns in imperatives of the *prends-le* type; cf. chapter four, §2.

characterize verse far more than they do prose. Cross-linguistically, verse forms make use of fundamentally oral features: rhyme, syllable count, position of stressed syllables, position of heavy syllables, position of tones, alliteration, etc. (cf. Gasparov, 1996). Because of their stronger links to the spoken form, the present study prioritizes the use of verse texts.

The earliest textual records from the *langue d'oil* area (modern-day northern France and Wallonia) date from the 9th century. Yet very few texts survive which were composed before 1100, and only three short texts date from the 9th and 10th centuries.[11] In fact, it is not until the second half of the 12th century that extensive textual evidence is available. However, from the end of the 12th century onwards, the historian of French is fortunate: a great variety of texts are preserved and are available in modern printed editions. For the present study, the earliest texts from which we can take any useful evidence date from the early 11th century: the three earliest texts are too short, only one is in verse (the *Sequence of Saint Eulalia*) and the precise form of the verse is unclear (cf. Ayres-Bennett, 1996: 33–34). Until the end of the 12th century, however, the choice of texts is still very limited. While most of the earliest texts are in verse, the use of texts from different dialects and genres, some preserved only in manuscripts copied over a century after they were composed, introduces a number of possible sources of linguistic variation which it would be more desirable to eliminate or control in a systematic manner. This is only possible from the end of the 12th century onwards, when the textual record is much richer. A detailed discussion of the corpus of texts on which the present study is based is given in chapter two (§1).

Since stress is not directly marked in the texts, aspects of the stress system must be inferred from lexical, phonological and (in the case of the stress group) syntactic properties of the texts. It is vital to be consistent in making these inferences. In the present study, this is achieved by automated processing of annotated source texts according to a fixed set of rules. The corpus used in this thesis is ground-breaking in combining syntactic and phonological annotation, and the majority of the studies in chapters three and four make use of quantitative data that cannot be obtained from any other corpus of French texts. As well as being essential for the work of the present thesis, a machine-processable corpus is a valuable resource for future study. The design of the corpus annotation scheme is documented in chapter two (§2).

11. Of these, only two are preserved in contemporary manuscripts. The earliest text (the *Strasbourg Oaths*) is preserved in a late 10th-century manuscript.

4 Structure of the thesis

The thesis is structured as follows. In chapter one, I focus specifically on the properties of group stress in modern French, and evaluate whether 'stress-conditioned' changes used as evidence for prosodic change in previous studies can be plausibly linked to the emergence of a group-stress system. Changes in the rhythm of verse are identified as the most promising source of data for dating the emergence of group stress. Chapter two documents the preparation of the corpus data for the studies in chapters three and four, including the texts used and the annotation scheme. Chapter three presents an extensive investigation of changes in the use of rhythm in MedFr versification. In addition to a primary focus on the chronology of the development, I also examine the effects of dialect and of text type on the use of rhythm. In chapter four, I focus on the interaction of stress and the phonological phrase in MedFr. Firstly, I investigate the stress pattern of the phonological phrase as predicted by both a word-stress and a group-stress grammar, showing that there is considerable overlap between the two, and consequently scope for reanalysis of the stress system. Secondly, I consider morpho-syntactic phenomena which have been linked in previous studies to the emergence of group stress, and demonstrate that the changed stress pattern of the phonological phrase is likely to have had an effect on the verb-second system of MedFr.

Chapter 1

Group stress: present and past

The present study takes the emergence of group stress as the most important aspect of prosodic change in the medieval period. The starting point for this study, therefore, must be to clarify the term 'group stress'. In section 1, I discuss the features of the modern French (ModFr) group-stress system, focusing in particular on the factors which determine stress placement. In section 2, a perceptual correlate of the group-stress system is discussed, the so-called 'stress deafness' effect first proposed by Peperkamp et al. (1999).

By 'beginning at the end' of the change with a clear characterization of the ModFr stress system, I hope to avoid the pitfalls of previous studies which, through less precise definitions of prosodic change, reach widely different conclusions both as to the dating of prosodic change and its interactions with other aspects of the linguistic system. In section 3, I examine some of the phonological and morpho-syntactic changes identified in previous studies as linked to prosodic change. Where the developments can be linked to a shift from word- to group-stress, I aim to date the changes to inform our understanding of the chronology of the prosodic development. Finally, in section 4, I consider evidence from the development of verse. I suggest that verse evidence is potentially a very powerful tool for investigating the changing prosody of MedFr, since the rhythm of verse provides an insight into how stress was perceived by medieval speakers.

1 Group stress in modern French

The term 'stress' in the present thesis denotes '[a] phonological feature by which a syllable is heard as more prominent than others' (Matthews, 2007:

383). [1] Stress is a 'phonological feature', distinct from its precise phonetic realization in terms of pitch, duration and intensity. However, for a syllable to be stressed, it must 'be heard as prominent'; thus some phonetic correlate of stress must be present in the speech signal. Gussenhoven (2004: 14–22) outlines the principal characteristics of this 'prominence' cross-linguistically. Phonetically, stressed syllables are treated with greater articulatory care, and vowels in unstressed syllables are likely to be shorter, more centralized and unrounded than their stressed counterparts. The phonetic opposition can also be phonologized, with a wider range of vowel phonemes present in stressed syllables, in particular phonologically long vowels. Finally, stressed syllables are associated with pitch movements. As phonemic tone is not present at any stage in the history of French, pitch movements are only associated with intonation contours.

Regular primary stress in modern French is fixed on the final syllable of a group of words (the stress group):

(1) [Mes amis] [vont arriver] [dans deux minutes]

The above example is divided into three stress groups, and the final syllable of each is stressed. The most salient phonetic features of stress in ModFr are a lengthening of the vowel and the presence of a pitch accent, most usually a rise (Lacheret-Dujour and Beaugendre, 1999: 41).

In later studies, I will attempt to 'reconstruct' the position of stress in medieval texts. For this reason, theories which make clear predictions about the position of primary and secondary stress and the domain of the stress group from phonological, lexical and syntactic features of the text are of particular interest. To this end, I begin with approaches which consider the stress group to be a syntactically delimited constituent on the prosodic hierarchy. Depending

1. 'Stress' is often distinguished from 'accent', but the criteria are not consistent. Beckman (1986) re-examines the perceived distinction between languages with 'stress accent' and those with 'non-stress accent'. A 'stress accent' system is argued to use 'to a greater extent material other than pitch' than a 'non-stress accent' system to mark accent (1986: 1). For Beckman, accent is associated with prominence, and also has an organizational function, dividing utterances into smaller 'accentual phrases' (1986: 1–2). Beckman's notion of 'accent' is thus similar to the notion of 'stress' adopted here, and Beckman's 'stress' denotes a particular type of phonetic realization of accent. Gussenhoven (2004) draws a different distinction. 'Accent' is directly associated with pitch, either lexical or intonational, serving as 'a marker for the insertion of a tone' (2004: 47). 'Stress', on the other hand, is defined in purely formal terms as a property of the foot: a prosodic constituent consisting usually of two syllables of which one is marked as 'more prominent' (i.e. more stressed) than the other. (For a detailed presentation of a metrical theory of stress, see Hayes (1995); Dell (1984) offers an analysis of ModFr stress in a similar framework.) Both Post (2000) and Gussenhoven (2004) analyse the prosody of ModFr in terms of the position of pitch accents rather the metrical concept of 'stress'. In the present thesis, unless discussing work which prefers the term 'accent', I will use only the term 'stress' in its non-formalist sense as defined above.

on how adequate these approaches prove to be empirically, I hope to ascertain the extent to which ModFr stress is determined by wider linguistic factors.

A second issue is the relationship between stress, the word and the stress group in ModFr. Lacheret-Dujour and Beaugendre (1999) argue that two fundamental positions have been adopted by previous studies:

(I) Words have no stress, and the position of stress depends purely on the limits of the stress group (e.g. Grammont, 1930; Pulgrum, 1965; Mertens, 1993).

(II) Content words have an underlying stress, but the stress is not always realized, being subject to a number of deaccenting rules. Stress groups are formed around those word-final stresses that are realized (e.g. Rossi, 1980, 1985).

From our point of view, the conceptual difference is important. Position I assumes that stress in ModFr is fundamentally a property of the stress group. This would represent a pure 'group-stress' system, in which the stress rule assigns primary stress to a higher-level prosodic constituent than the word. Position II, on the other hand, analyses ModFr as a fixed word-stress system with a number of deaccenting rules. Position I has more currency in the historical linguistic literature. Marchello-Nizia (1995: 174) highlights the loss of the 'autonomie accentuelle' of the word in French, and the prosodic autonomy of the word is also compromised by the loss of distinctive word boundaries. Pulgrum (1965) argues that in ModFr 'the lexical word in the context of an utterance has absolutely no phonological reality' (1965: 138), a view taken up by Klausenburger (1970) in his study of the diachronic development of French prosody. We will consider the data of ModFr in the light of these two contrasting views throughout the present discussion.

1.1 The syntax–prosody mapping: prosody dependent on syntax

1.1.1 Overview

The theory of prosodic phonology, which stems originally from work by Selkirk (1981), proposes a hierarchy of prosodic constituents which are defined using morphological and syntactic criteria (see, for example Nespor and Vogel, 1982, 1986; Selkirk, 1984, 1986, 1996; Selkirk and Shen, 1990). The prosodic hierarchy assumed by Selkirk (1996), and which I adopt in this thesis, contains the following constituents:

The Prosodic Hierarchy

utterance > intonational phrase > phonological phrase (PhPh) > prosodic word (PWd) > foot > syllable

In this framework, the 'stress group' of ModFr has been argued to correspond to the PhPh (e.g. Selkirk, 1986; Post, 2000); i.e., the constituent which immediately dominates the PWd. [2]

A core claim in work on the prosodic hierarchy is that prosodic phenomena are dependent on syntactically defined prosodic constituents. Nespor and Vogel (1982: 228; 1986: 168) define the PhPh in syntactic terms. Nespor and Vogel's definition being rather abstract, I reproduce Post's (2000) formulation of the rule:

Phonological Phrase formation rule

A Phonological Phrase groups together a lexical head (X) with all the items on its non-recursive side (i.e. the left) within the maximal projection and with any other non-lexical item on the same side.

(Post, 2000: 34)

Nespor and Vogel (1986: 169) assume that only nouns, verbs or adjectives are lexical heads. The PhPh formation rule is posited as a linguistic universal, and is exemplified by Nespor and Vogel with data from Italian, English, French and the Bantu language Chimwi:ni. The core prediction of this rule is that any constituent preceding a lexical head but within its maximal syntactic projection will fall within the same PhPh. In the syntactic theory adopted (cf. Jackendoff, 1977), this primarily affects adjectives before the noun, and pronouns and auxiliaries before the lexical verb.

An identical domain is identified by Selkirk (1986: 395–97) in ModFr, which she calls the 'small phonological phrase'. Selkirk offers the following example illustrating how phonological phrasing is derived from the syntactic structure of French:

(2)
$$
\begin{array}{lllllll}
 & {}_{S}[[\text{Sais-tu}]_{V} & & [[\text{quand}]_{Comp} & [\text{ils} & [\text{inviteront}]_{V} & \\
 & X^0] & & & & X^0] & \\
 & (\quad) & (& & &) & \\
 & [\text{un} & [\text{autre}]_{AP} & [\text{grand}]_{AP} & [\text{artiste}]_{N}] & {}_{NP}]_{VP}]] & \\
 & & & & X^0] & & \\
 & (& & &) & &
\end{array}
$$

(Selkirk, 1986: 396)

2. Nespor and Vogel (1986) argue for an intervening constituent, the clitic group. I adopt instead the more economical position of Selkirk (1996), who argues that the behaviour of clitics is better accounted for by assuming that they are either grouped together with lexical items into PWds, or immediately dominated by the PhPh.

In this example, the top line shows the syntactic structure, the bottom line the prosodic constituents (marked by parentheses) and the middle line the position of the lexical head which links the two structures. Here, the head of the verb phrase *inviteront* is grouped together with all the words to its left within the verb phrase (*quand*, *ils*). The head of the noun phrase *artiste* is grouped together with all the words to its left within the noun phrase (*un*, *autre*, *grand*). Stress falls regularly, and only, on the final syllable of the PhPh.

One of the arguments in favour of motivating prosodic constituents such as the PhPh cross-linguistically is that they describe the domain of phonological rules (Nespor and Vogel, 1986: 58–59). In ModFr, the PhPh describes the domain of the final primary stress rule.[3] In a number of languages, the PhPh describes the domain of stress clash resolution. A basic characterization of this rule is that stress on the final syllable of a word may be deleted or realized earlier in the word when the following word is stress-initial. Nespor and Vogel (1979) exemplify this rule with data from northern Italian. The adverb *metà* ('mid' or 'half') is oxytonic. When not followed by a stress-initial word, it is realized with final stress:

(3) [metà canzone]$_{\text{PhPh}}$
'half song' (Nespor and Vogel, 1979: 469)

When followed by a stress-initial word, the stress is 'retracted' onto the first syllable:

3. It is also suggested by Verluyten (1982), Nespor and Vogel (1986) and Selkirk (1986) that the PhPh is the domain of *liaison* in ModFr. Purely syntactic approaches to liaison are problematic. Morin and Kaye (1982), for example, note that liaison in ModFr is now only possible with two underlying consonants (/t/ and /z/). In some cases, the liaison consonant has been reanalysed as a purely morphological marker, being prefixed to plural post-nominal adjectives even where there is no etymological motivation (e.g. *des chemins de fer z-anglais*; Morin and Kaye, 1982: 321). Moreover, Post's (2000: ch. 3) experimental study confirms that the PhPh does not describe the domain of liaison correctly.

Despite the importance attached by historical linguists such as Pope (1952), Marchello-Nizia (1995) and Klausenburger (1970) to the emergence of *liaison* as part of prosodic change in MedFr, it will not be further discussed in the present thesis. A crucial assumption of these linguists is that like the loss of regular word stress, sandhi phenomena such as *liaison* represent a loss of the prosodic identity of the word. Marchello-Nizia (1995: 174), for example, contrasts the word-stress stage in the history of French 'où la frontière entre les mots était discernable' with the subsequent group-stress stage 'où [...] les mots sont liés prosodiquement selon les règles "de liaison"' (cf. also Klausenburger, 1970: 86). Yet I remain sceptical of associating the loss of word stress with the emergence of *liaison* in this way, primarily because it implies that sandhi phenomena are somehow incompatible with word stress. As we will see below with the case of Italian phono-syntactic gemination, not only are sandhi phenomena attested in word-stress languages, but they can even be conditioned by the presence of word stress. This is not to say definitively that the emergence of liaison cannot be linked to the emergence of group stress, but the case has not yet been convincingly made.

(4) [metà torta]$_{\text{PhPh}}$
'half cake' (*ibid.*)

However, when an oxytone and a stress-initial word do not fall within the same PhPh, stress retraction does not take place: [4]

(5) [i bigné]$_{\text{PhPh}}$ [scottano]$_{\text{PhPh}}$
'the doughnuts are hot' (*ibid.*)

Thus, the phonological rule of clash resolution is claimed to take place within the PhPh. Clash resolution of this kind has been observed cross-linguistically, and rule-based analyses all follow the same basic principle: there is an underlying word stress which is retracted or deleted. [5] Using experimental data, Post (2000: 52–53) demonstrates that a very similar process takes place within the PhPh in ModFr. [6] In particular, the final stress on a polysyllabic adjective is perceived to fall on the initial syllable when it is followed by a monosyllabic head noun:

(6) [de jolis airs]$_{\text{PhPh}}$ (Post, 2000: 37)

This type of clash resolution occurs in 143 of 162 tokens of adjective–monosyllabic noun combinations in Post's corpus, and there are only six cases in which it has clearly failed to apply (Post, 2000: 53).

To summarize so far, the theory of the PhPh emphasizes the dependence of prosody on syntax. Prosodic constituents are defined on the basis of syntactic criteria. Their boundaries are determined by (i) the lexical category of a word and (ii) whether or not it is the head of a maximal projection. Phonological rules, such as clash resolution and the position of primary stress, are dependent on prosodic constituents. Therefore, we may predict that prosody depends on syntax. With regard to the two views of the relationship between primary stress and the stress group, theories of the syntax–phonology mapping are broadly of type (II), assuming that each word has an underlyingly present primary stress.

1.1.2 The problem of phonological phrase restructuring

The first case which proves cross-linguistically problematic for the purely syntactic approach is the treatment of complements which follow the lexical

4. Examples are presented in the framework of Nespor and Vogel (1986), rather than Nespor and Vogel (1979).

5. On clash resolution, see Liberman and Prince (1977), Selkirk (1984), Hayes (1984, 1995), Nespor and Vogel (1979, 1989), and Gussenhoven (1991). On stress deletion in clash resolution, see chapter four, section 1.1.2.

6. Mazzola (1992) and Hoskins (1994) offer similar analyses of stress clash in French.

head of a maximal projection (for example, post-nominal adjectives). Nespor and Vogel (1986: 172–73) propose an optional rule of PhPh restructuring, in which non-branching post-head complements (i.e. single words) within the maximal projection may be incorporated into the PhPh. Evidence for such a rule is found in standard Italian. Unlike in northern Italian, here clash resolution does not take place; instead, vowel-final oxytones cause the initial consonant of the following word to be lengthened, a phenomenon known as *raddoppiamento fonosintattico* ('phono-syntactic gemination'). Again, this rule is argued to apply only within the PhPh:

(7) [Ho **vv**isto]$_{\text{PhPh}}$ [tre **cc**olibrí]$_{\text{PhPh}}$ [molto scuri]$_{\text{PhPh}}$
'I saw three very dark hummingbirds' (Nespor and Vogel, 1986: 171).

In this example, initial consonant lengthening applies to *visto* and *colibrí*, but not to *molto*, since the preceding vowel-final oxytone is in a separate PhPh. However, between a head and a non-branching post-verbal complement (e.g. noun–adjective sequences), consonant lengthening is optional:

(8) [I caribú]$_{\text{PhPh}}$ [nani]$_{\text{PhPh}}$ [sono estinti]$_{\text{PhPh}}$

(9) [I caribú **nn**ani]$_{\text{PhPh}'}$ [sono estinti]$_{\text{PhPh}}$
'Dwarf caribous are extinct'

(Nespor and Vogel, 1986: 172–73)

In the second of the two cases, the two PhPhs are argued to have been restructured into a single PhPh′. Consonant lengthening is therefore attested. However, since both (8) and (9) are possible variants, we conclude that the restructuring rule is not obligatory.[7]

An optional rule of this kind weakens the foundation of the theory that prosody depends only on syntax (as noted, for instance, by Jun, 1998: 190–91), and in so doing weakens the predictive power of the approach. However, the predictive strength of the theory can be retained, and improved, if non-syntactic factors are included in the model. Post (2000) argues that syllabic length conditions the likelihood of PhPh restructuring in ModFr. She reports that restructuring takes place in 73% of cases in her corpus where a noun is followed by a monosyllabic adjective (2000: 84):

(10) [des enfants sages]$_{\text{PhPh}'}$ (Post, 2000: 84)

Post demonstrates that the PhPh has been restructured using two pieces of evidence. Firstly, the final vowel of a PhPh is significantly lengthened compared

7. Nespor and Vogel (1986: 173) suggest that restructuring may be constrained by non-syntactic factors such as speech rate, with restructuring more common in fast speech styles.

with vowels within the PhPh (2000: 51–52). In tokens such as (10), the final vowel of *enfants* is not lengthened. Secondly, clash resolution applies. On the other hand, where the noun is followed by a longer complement, restructuring fails to take place in the majority of cases (63%) (Post, 2000: 84):

(11) [des enfants]$_{\text{PhPh}}$ [intelligents]$_{\text{PhPh}}$

Post models this alternation using a constraint-based Optimality Theory framework. Such an approach has advantages over a rule-based model, since minor changes in the ranking of constraints correctly predict attested variants. However, as the formalism of the approach is not important to the present thesis, I summarize the predictions of the default constraint ranking as follows: [8]

> **Post's (2000) PhPh formation rule**
> A PhPh is preferentially a full XP; [9] however, if a word of more than one syllable follows the head of the XP, the XP is split into two PhPhs after the head.

A similar combination of syntactic and length-based rules are adopted by Verluyten (1982) and Delais-Roussarie (1995). Verluyten (1982), following an initially syntactic definition of the *groupe accentuel* (i.e. the stress group of the present discussion) proposes a number of rhythmic readjustment rules. A short *groupe accentuel* is combined with the neighbouring group to which it has the strongest syntactic relation (Lacheret-Dujour and Beaugendre, 1999: 129). Delais-Roussarie (1995), using a constraint-based approach similar to that of Post (2000), argues that two rhythmic principles, one limiting the maximum length of the PhPh to six syllables, and one favouring *groupes accentuels* of similar length, both outrank the syntactic constraint requiring a boundary between *groupes accentuels* after a lexical head (Lacheret-Dujour and Beaugendre, 1999: 150–51).

In summary, both rhythmic and syntactic criteria play a role in establishing the extent of the stress group in ModFr. The models discussed have retained a

8. In the constraint-based Optimality Theory framework adopted by Post (2000), three ranked constraints govern the formation of the PhPh. Highest ranked is a length constraint PhPhLength(1), which is violated if more than one syllable follows the head of an XP within the same PhPh (2000: 91). Below this, an alignment constraint AlignXP requires the right edge of a PhPh to occur always and only at the right edge of an XP (2000: 89). Below this, a second alignment constraint AlignX′ requires the right edge of a PhPh to occur always and only at the right edge of a lexical head (2000: 87–88). The result of the default ranking of these constraints is that an XP may only be split into two PhPhs (violating AlignXP) if not splitting the XP results in a long complement following the lexical head (violating the higher-ranked PhPhLength(1)). Where necessary, splits are always made after the lexical head (avoiding a violation of AlignX′).

9. I.e. a syntactic phrase with a lexical head: noun phrase, adjective phrase or verb phrase.

deterministic view of the stress group: given a particular syntactic structure, and a particular number of syllables in the lexical items concerned, the domain of the stress group is predictable. Yet a closer look at the ModFr data suggests that the formation of the stress group may not always be so regular.

1.1.3 The problem of variation

It is clear from Post's study that there is variation in PhPh restructuring: we saw above that it occurs in only 73% of cases in which a noun is followed by a monosyllabic adjective. Other studies claim that variation is more widespread even than this. Fónagy (1980) examines a number of lexical, syntactic, diamesic and idiosyncratic factors affecting the realization of stress in French, arguing that French has an 'accent probabilitaire'. Fónagy's view is widely shared: Lacheret-Dujour and Beaugendre, for example, argue that 'la forme accentuelle [...] n'existe pas en tant que structure fondamentale et invariante, elle est construite par les sujets' (1999: 40).

Post, who presents a deterministic model of PhPh formation (2000: ch. 4), notes that such a model does not account for all the tokens attested in her corpus. In particular, she notes the following areas of variation:

(i) No restructuring of monosyllabic adjective with preceding noun (2000: 101):

(12) [des en<u>fants</u>]$_{\text{PhPh}}$ [<u>sages</u>]$_{\text{PhPh}}$ (27% of cases)

(ii) Partial or complete restructuring of syntactically branching post-nominal complements (2000: 61):

(13) [des hivers <u>autres</u>]$_{\text{PhPh}}$ [qu'en A<u>frique</u>]$_{\text{PhPh}'}$ (25% of cases)

(14) [des hivers autres qu'en A<u>frique</u>]$_{\text{PhPh}'}$ (10% of cases)

In (13), only the monosyllabic head of the branching adjectival complement *autres qu'en Afrique* has undergone restructuring with the previous PhPh. In (14), the entire branching complement has been restructured. Post shows that simple modifications to her constraint-based model can produce the variant tokens (2000: 102–3) and suggests that the variation may be conditioned by 'factors such as boundary weight and rate of speech' (2000: 107).[10] While this demonstrates the strength of the constraint-based formalism, introducing variation conditioned by performance factors into the model fundamentally

10. However, the variant *[des hivers autres] [qu'en Afrique]* is not discussed; on my understanding, this form cannot be selected by the constraints in the model, whatever the ranking.

alters the nature of the analysis: stress position can no longer be determined solely by language-internal factors.

1.2 Intonation-based approaches

The approaches to stress in ModFr discussed to this point have two important points in common. Firstly, syntactic (and rhythmic) factors are assumed to govern the domain of the stress group, and hence the position of primary stress. Secondly, the rhythmic concept of 'stress' is distinguished from the intonational concept of 'tone', even though, in the surface form, stress is realized by a pitch movement (cf. Post, 2000: 109).

An alternative approach is to consider the stress group as a unit of intonation. For Jun and Fougeron (1995, 2000, 2002), working in the Autosegmental-Metrical framework, the stress group is an Accentual Phrase, defined as 'the lowest *tonal unit* in French' [my emphasis] (Jun and Fougeron, 2000: 210). The accentual phrase has a fixed tonal pattern, which may be defined as /LHiLH*/: i.e. a low tone (L) followed by an 'initial' or 'secondary' high tone (Hi), followed by a low tone, followed by a high tone (H) 'associated to a stressed syllable' (*) (Jun and Fougeron, 2000: 211). This tonal pattern is phonologically underlying, and may be variably realized phonetically, depending on the length of the accentual phrase (cf. Jun and Fougeron, 2002). In long accentual phrases (i.e. more than four syllables; Jun and Fougeron, 2002: 156), the full four tones are realized:

(15)

L	Hi					L	H*
le	dés-	a-	gré-	able		gar-	çon

(Jun and Fougeron, 2000: 215)

Note that the initial [Hi] tone is associated with the first syllable of the first content word rather than the initial clitic.[11] In short accentual phrases, the phonetic realization of the pattern may simply be [LH*]:

(16)

L	H*
Ma-	rie

(Jun and Fougeron, 2000: 215)

Whatever the descriptive accuracy of Jun and Fougeron's model,[12] what is interesting from our point of view is that it represents a totally inverse conception of the nature of group stress. Firstly, the boundaries of the accentual phrase are not derived from syntactic and rhythmic principles, they are

11. This is not an exceptionless generalization: see example (23) below.
12. See Post (2000: 176–86) for a detailed critique of many aspects of the proposal.

observed from the phonetic form of the utterance (Jun, 1998: 190). The approach thus offers a descriptive rather than a deterministic approach to the stress group. Secondly, it is not clear whether the authors consider underlying stress separately from the tonal realization of the accentual phrase (i.e., whether they adopt position I or II above). Indeed, they suggest that the final /LH*/ tone has a 'double association': 'LH* marks the right edge of an AP [accentual phrase], but is also associated with the stressed syllable of an AP, i.e., the final full syllable of the last content word [sic.] of an AP' (2002: 151). If /LH/ is associated with the right edge of an accentual phrase, this implies position I: stress is a pitch movement associated with a prosodic constituent edge. If /LH/ is associated with a stressed syllable, this implies position II: stress is a property of the word, realized by an associated tonal movement. The authors offer no further criteria to determine the position of stressed syllables. Working in a similar framework, Féry adopts a position I approach more unequivocally, arguing that 'the tonal excursions associated with phrase boundaries, both at the beginning and the end of phrases, are boundary tones' (2001: 177).

Where syntactic factors are used in intonation-based models, it is rarely to define the stress group. Di Cristo and Hirst (1993, 1997; Di Cristo, 1999, 2000) also working within an intonational framework, define the highest constituent in their model, the *unité intonatif* (UI), as having boundaries determined by the edges of a syntactic constituent:

(17) [Il a rencontré les écrivains de la contestation]$_{\text{UI}}$ [lorsqu'il a fait son reportage]$_{\text{UI}}$

(Di Cristo, 2000: 35)

Here, the edge of a *unité intonative* corresponds to the right edge both of a noun phrase and a verb phrase. The *unité intonative*, however, is not the domain of primary stress. Nor is the *unité tonale*, the minimal unit of intonation containing the tonemes /LH/. The right edge of an *unité tonale* is aligned with both the first and the final syllables of any lexical word (Di Cristo, 2000: 34):

(18) [(Il a ren)(contré) (les é)(crivains) (de la con)(testation)]

(Di Cristo, 2000: 35)

The unit corresponding to the stress group, the *unité rythmique*, falls in between the *unité tonale* and the *unité intonative* on the prosodic hierarchy proposed by Di Cristo and Hirst (1997). Like the accentual phrase of Jun and Fougeron, and unlike the more deterministic rules used for the *unité tonale* and the *unité intonative*, this unit is descriptive: it is defined as 'ayant pour

tête un accent primaire' [emphasis original] (Di Cristo and Hirst, 1997). The means by which *unités tonales* are grouped into *unités rythmiques* are variable, and depend on rhythmic constraints, speech style and individual strategies (Di Cristo and Hirst, 1997: 91). The following constraint on the formation of the *unité rythmique* illustrates the fundamentally non-deterministic nature of the approach:

> Une unité rythmique est Unaire, Binaire ou Illimitée.
>
> (Di Cristo and Hirst, 1997: 93)

In other words, it contains one, two, or more than two *unité tonales*: a generalization with no predictive power whatsoever.

The existence of intonation-based approaches to group stress highlights the degree to which the distinction between intonation and stress is not clear-cut in ModFr. Ladd (2008) argues that intonation is a post-lexical feature, carrying sentence-level meanings. Intonation interacts with, but does not include, 'features of stress, accent and tone that are determined in the lexicon, which serve to distinguish one word from another' (2008: 6). Group stress is not easy to categorize from this definition. As we will see in section 2, it is not clear that it is 'determined in the lexicon', and it certainly does not 'distinguish one word from another'. However, neither does the tonal pattern of the stress group function as a bearer of sentence-level meanings.

To summarize, intonation-based approaches to the stress group in ModFr are generally descriptive rather than deterministic with regard to the placement of primary stress. In one sense, they are of little use to the historical linguist, since in order to use these theories, an acoustic signal is required to derive the position of the phrase boundaries. In another sense, the fact that ModFr stress can be represented in this way highlights the parallels between group stress and intonation. Moreover, the existence of analyses which take the position of stress as the best cue for delimiting prosodic constituents only serves to emphasize the difficulties encountered by deterministic, syntactico-rhythmic approaches with regard to the variation in phonological phrasing in ModFr. Finally, intonation-based approaches move us more towards a position I approach to the relationship between the stress group and primary stress. Thus, for Jun and Fougeron (1995, 2000, 2002), the tonemes which realize primary stress are associated with the edge of a prosodic constituent. Moreover, as we will see, the secondary /LHi/ tonemes are not associated with an underlyingly stressed syllable at all.

1.3 The problem of secondary stress

In the overwhelming majority of cases, group-final primary stress falls on the final syllable of a content word.[13] As such, the data is compatible either with a word-final stress analysis or a group-final stress analysis (positions I or II above). However, within the stress group, secondary stressed syllables may also be realized.[14] The position of these secondary stresses within the the stress group should, in principle, allow us to distinguish between positions I and II, since a position II approach predicts that secondary stress will normally occur word-finally within the stress group.

A strong argument in favour of a position I approach to secondary stress is the observation that secondary stress falls on the first syllable of the first content word of a stress group (Fónagy, 1980; Di Cristo and Hirst, 1997; Jun and Fougeron, 2000). Fónagy (1980) considers stress groups to be marked by a single *arc accentuel*: a tendency to realize a stress on both the initial and final syllables of stress groups. He cites the following patterns as illustrative:

(19) à protéger la France

(20) la majeure partie

(21) les hautes pressions (Fónagy, 1980: 143)

The group-initial stress is most often described as *rhythmic* in nature (Di Cristo and Hirst, 1997: 82–83; Lacheret-Dujour and Beaugendre, 1999: 44); Di Cristo and Hirst (1997) deem it a 'secondary' stress. For Jun and Fougeron (1995, 2000, 2002), secondary stress is characterized by the initial /LHi/ pitch movement in the accentual phrase. They argue that the /Hi/ tone is (i) optional, (ii) 'weaker than H* in its duration and pitch' (2000: 213), and (iii) generally realized on the first syllable of a content word. The phrase-initial stress, by its optionality and its weaker phonetic nature, is secondary to the phrase-final stress.

Unlike primary stress, the realization or non-realization of secondary stress is not determined by lexical or even syntactic factors, but by a variety of

13. The only exception is the stressed realization of post-verbal pronouns in imperatives such as *prends-le*. This can by reconciled with a word-stress approach either by assuming that these stressed pronominal forms are included separately in the lexicon alongside their unstressed counterparts (Gussenhoven, 2004: 265), or by arguing that all group-final lexical items must be parsed as prosodic words (Selkirk, 1996; Post, 2000; and chapter four, §2.3).

14. Group-initial secondary stress is assumed to be unemphatic and distinct from the *accent d'insistance* (cf. Di Cristo and Hirst, 1997: 81–84). While I will argue in chapter four (§1) that initial secondary stress is medieval in origin, the status and origins of the *accent d'insistance* are less clear (Di Cristo, 1999: 166–69), and I will not consider it further in this thesis.

rhythmic performance constraints. [15] In Post's corpus (2000: 101), 9% of tokens of adjective–monosyllabic noun combinations show no secondary stress:

(22) de jolis airs

A further argument in favour of the position I approach is that in a long sequence of clitics, a secondary stress may be realized on a clitic:

(23)

il	faut	que	je	le	lui	en	donne
[L	H*]	[L	Hi			L	H*]

(Jun and Fougeron, 2002: 164)

However, word-initial stress is not incompatible with a position II approach to secondary stress. Post (2000), for instance suggests that secondary pitch accents are preferred on the final syllable of words even within the stress group:

(24) de jolis enfants (Post, 2000: 96)

However, the accent may be realized initially in cases of stress clash:

(25) de jolis airs (Post, 2000: 98)

Post's analysis assumes that secondary accents are assigned to PWds within stress groups and shifted (at least perceptually) to the first syllable only in cases of stress clash. [16] However, in a discussion of common prosodic variants, Post notes that patterns such as *de jolis enfants* are attested in her corpus, but does not provide an indication as to their relative frequency. Such patterns cannot be modelled using a word-based approach to secondary stress. [17] Indeed, in an analysis similar to that of Post, Gussenhoven (2004: 260) prefers to account for such patterns with a constraint requiring left-alignment of a pitch accent with a PhPh boundary. Such an analysis reinforces a position I view

15. E.g. the number of syllables in the group, speech rate, speech style, location of the group in the utterance, adjacent tones, and the individual speaker (Jun and Fougeron, 2002: 154).

16. I have expressed the analysis in rule-based terms; Post uses a constraint-based approach to identical effect.

17. Post (2000) suggests the following solution, which I feel is problematic. The position of secondary stress is regulated by four constraints: NoClash, prohibiting a stress clash; Hammock, requiring each PhPh to have a second stress; RightmostPWd, requiring each prosodic word to have a final stress, and LeftmostPWd, requiring each PWd to have an initial stress. Post (2000: 104) argues that if word-initial stress is preferred to word-final stress, patterns such as *de jolis enfants* are correctly generated (note that the position of primary stress is fixed by higher-ranked constraint and is therefore unaffected). A difficulty arises with patterns containing three PWds, such as *une assez jolie jupe*. Post (2000: 104) rules out *une assez jolie jupe* (word-final instead of word-initial stress) and *une assez jolie jupe* (more than one secondary stress, violation of Hammock). The form *une assez jolie jupe* is thus argued to be the optimal candidate (2000: 104). However, the form *une assez jolie jupe*, not present in Post's analysis, seems to me to be co-optimal. Since such a form is not common in ModFr this is an incorrect prediction of the analysis. Gussenhoven (2004: 260) makes a similar criticism, noting the unlikelihood of forms such as ??*des personnes* **a***litées*.

of stress in ModFr, in which both primary and secondary stress are assigned to the edges of the stress group, and in which the PWd is irrelevant for stress assignment.

However, this 'extreme' position I view of French fails to account for all the data. Di Cristo (1999: 164–65) questions the view that content words within a stress group show no distinct word final stress. Gussenhoven (2004: 258) argues that in formal speech styles, a PhPh may contain three pitch accents, and if so, the medial accent is invariably PWd-final:

(26) très jo**lis** garçons

(27) exce**llent** travail (Gussenhoven, 2004: 260)

Here, a medial accent is realized on the final syllable of the pre-nominal adjective. Based on Gussenhoven's examples, it could be objected that the medial accent is realized here not because the syllable is word-final but rather because it is the only position in which a medial stress can occur without clashing. Moreover, rhythmic factors may play a role in the realization of secondary stress: Delais-Roussarie (1995), for example, proposes a (violable) NO-LAPSE constraint, restricting sequences of more than two unstressed syllables. Pasdeloup also argues that secondary stress has a rhythmic 'regulatory function' (1992: 339). She argues that it is most frequently realized on the first consonant-initial syllable of a word, or 'on the antepenultimate syllable of a word ending with primary stress' (1992: 338), an observation also made by Garde (1968) and Verluyten (1984) (Lacheret-Dujour and Beaugendre, 1999: 44). However, Di Cristo and Hirst (1997: 82) note that given the relative rarity of monomorphemic words containing more than three syllables in French, the antepenult is often the initial syllable. Pasdeloup also suggests that the position of secondary stress is dependent on morphemic structure (Pasdeloup, 1992: 339), being commonly realized on a morpheme boundary in a polymorphemic word (e.g. *mé**lo**dramatique*) (1992: 339), an observation which is difficult to account for in a position I analysis of secondary stress.

It is thus difficult to defend a 'pure' position I analysis, in which secondary stress simply marks the left edge of a group. Sensitivity to morphological structure, including the distinction between content and function words, demonstrates that the PWd is still relevant for secondary stress assignment. However, even allowing for clash resolution, a position II analysis fails to account for the frequent realization of secondary stress in word-initial position. Gussenhoven argues that pitch accent position in ModFr is determined by 'a complex interaction of constraints that locate pitch accents at constituent boundaries', the

relevant prosodic constituents being both the PhPh and the PWd (2004: 265). Crucially, however, accent assignment is 'entirely postlexical' (2004: 275): no syllable is specified as stressed in the lexicon.

1.4 Summary

I return to the core questions posed at the beginning of this section. We have seen that attempts to define the stress group deterministically based on syntactic and rhythmic criteria are relatively successful. However, variability not conditioned by language-internal factors cannot be modelled in this way. We have also seen that stress placement can be described in terms of a higher-level prosodic constituent than the PWd. The PhPh, a unit which may be defined independently of stress on the basis of syntactic constituency and syllabic length, is characterized by a regular primary stress on the final syllable. The facts of secondary stress make it difficult to defend the view that there is an 'underlying' word stress independent of phrasing structure (position II): instead, both primary and secondary stress are assigned postlexically based on the position of PhPh boundaries and, to a lesser extent, PWd boundaries. It is this combination of postlexical stress assignment and the fact that stress position is dependent on a higher-level prosodic constituent than the PWd which form the key features of the ModFr group-stress system.

2 Stress 'deafness' in modern French

In the previous section, the properties of ModFr group stress were considered from the point of view of the position of stressed syllables. In this section, I will consider how and why speakers' perception of stressed syllables differs in a fixed-stress language such as French from that of speakers of a free-stress language, such as Spanish.

As primary stress in ModFr is fixed on the final syllable of the group, there are no lexical items distinguished only by the position of stress. Work by Dupoux, Peperkamp and colleagues (Dupoux et al., 1997, 2001, 2008, 2010; Peperkamp et al., 1999) shows that when native French speakers are presented with minimal pairs distinguished only by the position of the stressed syllable, they experience difficulty in distinguishing the stimuli. The authors name this phenomenon 'stress deafness'. They go on to argue (Dupoux and Peperkamp, 2002; Peperkamp and Dupoux, 2002; Peperkamp, 2004) that stress deafness is a permanent consequence of acquiring a first language such as French where

stress position is not distinctive. They establish a typology of stress systems which are predicted to cause stress deafness and test these predictions theoretically and experimentally.

The existence of stress deafness is demonstrated by a minimal pair discrimination task. Peperkamp et al. (1999) and Dupoux et al. (2001) report experiments in which subjects were given a pair of nonce stimuli, and told to associate each word with a separate button. Subjects were tested with pairs of stimuli distinguished by the position of stress, (e.g. /ˈpiki/ vs. /piˈki/), and, additionally, in a control condition using a segmental distinction (e.g. /ˈkupi/ vs. /ˈkuti/). A block containing a sequence of two to six tokens of these stimuli was played to subjects. Subjects were then asked to recall the sequence by pressing the correct sequence of buttons for the tokens heard. French speakers consistently showed a far higher error rate in the stress condition than in the control condition. For example, Peperkamp et al. (1999) report an error rate of over 60% for French native speakers in the stress condition with blocks of four or more tokens, compared with a 20–40% error rate in the equivalent segmental condition.

Stress deafness effects of this kind are linked to the nature of stress in the speaker's first language. In an identical minimal pair discrimination task, native speakers of Spanish (a free-stress language with stress-distinguished minimal pairs) performed equally well in the stress condition and the control condition (Peperkamp et al., 1999). Moreover, stress deafness effects are not eliminated by later acquisition of a free-stress language. For instance, it is shown in Dupoux et al. (2008) that even at an advanced stage of fluency, French late learners [18] of Spanish still perform significantly worse than native Spanish speakers in the stress condition. This is the case even where one of the stimuli is a genuine word in Spanish (e.g. *coraz<u>ón</u>* 'heart' vs. **co<u>ra</u>zon*). Stress deafness seems to be deeply rooted in the linguistic system of native French speakers, such that even near fluency in Spanish cannot eliminate the effect. [19] Dupoux et al. (2008) argue that stress deafness stems from an inability to encode stress features in the mental lexicon. This hypothesis implies that infants decide at an early stage of language acquistion whether or not stress is a lexically relevant feature. If not, stress is not included in representations of L1 lexemes, and the ability to encode stress in the lexicon is lost for life. However, it is not the case that all languages with fixed stress show a similar degree

18. I.e. those who began learning after the critical period of first language acquisition.

19. This is not to say that extensive exposure to and use of a free-stress language cannot reduce the strength of the effect, cf. Tremblay (2009) on Canadian French learners of English.

of stress deafness. Peperkamp and Dupoux (2002) test speakers of Finnish, Hungarian and Polish, all of which are fixed-stress languages. Stress deafness is not attested equally robustly among speakers of these languages. Finnish and Hungarian speakers exhibit a degree of stress deafness, but markedly less so than French speakers. Polish speakers, however, do not show a strong stress deafness effect.

Building on work in Dupoux and Peperkamp (2002), Peperkamp (2004) distinguishes languages in which stress is 'surface observable' from those in which it is not:

> A phonological regularity is *surface observable* if it can be inferred from bare utterances, that is, utterances in which word boundaries are not marked.
>
> (Peperkamp, 2004: 103)

Not all fixed-stress rules are surface observable. Those which are sensitive to morphological information are not, as knowledge of morpheme boundaries is required to detect the regularity (2004: 101). On the other hand, regular initial or final stress is always surface observable. Regular penultimate stress (such as found in Polish) may be surface observable, but only if the language contains no stressed monosyllabic words. Stressed monosyllabic words produce a surface alternation between penultimate stress (where the utterance ends in a word of more than one syllable) and final stress (where it ends in a monosyllable).[20] Before the regularity behind this surface alternation can be learnt, the acquirer must have additional, lexical knowledge (i.e. which content words are monosyllabic and which are polysyllabic). Since Polish contains stressed monosyllables of this kind, the regularity is not surface observable. Consequently, Peperkamp predicts that stress in languages such as Polish *will* be lexically encoded. This would explain the lack of stress deafness effects attested in Peperkamp and Dupoux (2002). Moreover, Peperkamp (2004) hypothesizes that languages in which stress is surface observable should not contain lexical items with irregular stress patterns, even loan words, as speakers are unable to encode such patterns. For example, English loan-words in French are all oxytonic (e.g. *tun<u>nel</u>*, *shop<u>ping</u>*). While on the basis of a wide-ranging typological survey, this does not prove to be an exceptionless universal, it holds true for 80% of languages with surface-observable stress (2004: 114).

To summarize, native speakers of languages with surface-observable stress

20. Peperkamp (2004) prefers the term 'subminimal word' rather than monosyllable, as penultimate stress may apply at the moraic rather than the syllabic level in some languages.

do not develop a phonological representation of stress in their mental lexicon. Therefore, they experience 'stress deafness': difficulty in perceiving stress-contrasted minimal pairs. The effect is linked to first language acquisition, as it remains even among second language learners of a language which does not show surface observable stress. The effect is a property of 'fixed-stress' languages, and previous work in the area does not specify whether the distinction between a fixed word stress and a fixed group stress has an impact on the strength of the effect. However, we note that of all the languages studied by Peperkamp and Dupoux (2002), it is modern French which shows the greatest stress deafness effects. Whether this simply a consequence of a surface-observable stress rule, or whether the effect is intensified by the fact that the stress-bearing unit is the stress group rather than the word is unclear. One thing is certain: both group stress and stress deafness presuppose that stress position is determined by a regular fixed-stress rule, and does not form phonologized lexical oppositions. Therefore, we predict that group stress cannot exist without stress deafness.

3 Stress-related phenomena in medieval French

In the introduction, we noted the problems inherent in studying an aspect of a language which is not present in the written form. Previous studies of prosodic change have tended to infer the properties of stress from apparently stress-related phonological and morpho-syntactic phenomena which are directly observable in the texts.

In this section, I will examine the insights provided by the study of stress-related phenomena. Firstly, I will consider evidence from segmental phonology (§3.1). The focus here will be to fix a *terminus a quo* for the emergence of group stress by establishing the latest point at which we have positive evidence for a word-stress system. Phonological evidence for the existence of group stress is more difficult to find, since in the ModFr group-stress system stress does not have a significant influence on segmental phonology.[21] Secondly, I will consider a proposed effect of stress on the morpho-syntax of MedFr: the existence and subsequent loss of a 'strong initial position'. Here, the main focus of the discussion will be to evaluate just how 'stress related' this phenomenon

21. A rare exception is vowel lengthening under primary stress, which takes place in closed syllables ending with a voiced fricative or approximant (Tranel, 1987: 49). Elsewhere, however, Tranel (1987: 33–35) emphasizes the lack of effect that stress has in ModFr in comparison with English.

actually is, and thus whether such evidence is valuable for a study of the emergence of group stress.

3.1 Evidence from phonological change

In the development of French from vulgar Latin, the position of stress had a major impact on the phonological development of words (Pope, 1952: §223). Two stress-conditioned changes in particular will be considered here: (i) reduction and loss of unstressed vowels and (ii) lengthening and diphthongization of primary stressed vowels.

One of the consequences of the loss of unstressed syllables was the emergence of a fixed-stress system in MedFr from the free stress of vulgar Latin. I begin in section 3.1.1 by trying to date the emergence of this rule, since a fixed-stress rule of this kind is a necessary condition for the emergence of stress deafness and group stress. The diphthongization of primary stressed vowels is then examined (§3.1.2). Since diphthongization affects the primary stressed syllables of all content words in all contexts, such a change shows that stress rules still apply to individual words. I also consider patterns of vowel reduction to schwa within words. Vowel reduction is typically a stress-conditioned change, and, like diphthongization, applies at the level of the word in MedFr. Taken together, these three changes will provide a *terminus a quo* for the emergence of group stress.

3.1.1 The emergence of fixed stress

Classical Latin had a syllable weight sensitive stress rule (Pope, 1952: §§211–13). If the penult was heavy (i.e. contained a coda consonant or a long vowel), it was stressed:

(28) PAR.TEM > vulgar Latin */ˈpar.tɛ/

(29) PER.SŌ.NA > vulgar Latin */pɛr.ˈso.na/

If the penult was light, stress fell on the antepenult:

(30) VĪ.VĔ.RE > vulgar Latin */ˈvi.vɛ.rɛ/

(31) AN.GĔ.LEM > vulgar Latin */ˈan.gɛ.lɛ/

The position of classical Latin stress was therefore predictable, although the rule was a complex one. In the evolution of classical Latin to vulgar Latin, phonemic vowel length distinctions were replaced by phonemic vowel quality distinctions. Thus vulgar Latin, like modern Spanish or Italian, had a free-stress system (cf. Väänänen, 1981: §§42–48). For example, there is nothing

in the segmental representation of *persona* in (29) to suggest that its stress pattern would differ from that of *vivere* in (30) or *angele* in (31).

Yet in MedFr, a fixed-stress system had re-emerged. The MedFr stress rule may be descriptively summarized below: [22]

MedFr stress rule

Stress the final syllable of the word, unless the vowel in that syllable is a schwa, in which case stress the penult.

The re-emergence of fixed stress was caused exclusively by the reduction and deletion of unstressed vowels. Indeed, the primary stress in MedFr fell on exactly the same syllable of the word as it had done in vulgar Latin (Pope, 1952: §215; Väänänen, 1981: §48). This is shown by the following examples:

(32) PAR.TEM > MedFr <part> /ˈpart/

(33) VAC.CA > MedFr <vache> /ˈva.tʃə/

(34) IN.FĔR.NUM > MedFr <enfer> /ɛ̃n.ˈfɛr/

(35) PER.SŌ.NA > MedFr <persone> /pɛr.ˈsõ.nə/

(36) VĪ.VE.RE > MedFr <vivre> /ˈvi.vrə/

Examples (32–36) show the unchanging position of primary stress in the development of French from Latin. Unlike ModFr group stress, MedFr stress was not regularly final, and did not become so until final schwa was lost from pronunciation. This change post-dates the medieval period, and while there is evidence that final schwa begins to be lost in some contexts in the 15th century, it is still attested by grammarians of the 16th century such as Palsgrave (cf. Pope, 1952: §271, §273).

A number of previous studies of prosodic change consider the re-emergence of a fixed-stress rule as the first stage of prosodic change in French (Klausenburger, 1970; Kukenheim, 1971; Banniard, 1998). From our point of view, it is clear that the emergence of a fixed-stress rule is a necessary precondition for the emergence of group stress and stress deafness effects. Following the work of Dupoux and Peperkamp on stress deafness (§2), we predict that only fixed-stress languages can show stress deafness effects. Despite the continued presence of final schwa, we may suppose that the MedFr stress rule was surface-observable. While there was surface alternation between penultimate and final stress, it was conditioned purely by phonetic factors (i.e. whether the final syl-

22. Cf. Pope (1952: §223): 'Old French [was] a language in which all words were *oxytone* or *paroxytone*, and, if paroxytone, then always ending in a syllable containing /ə/.'

lable was a reduced vowel or not).[23] Moreover, Palsgrave's evidence, cited in the introduction, strongly suggests that French had developed a group-stress rule by the 16th century, despite the continued presence of final schwa:

> **MedFr group-stress rule (hypothetical)**
> Stress the final syllable of the *stress group*, unless the vowel in that syllable is a schwa, in which case stress the penult.

I therefore assume that the MedFr fixed word-stress rule was a sufficient condition for the emergence of group stress.

Two changes caused the free stress of vulgar Latin to develop into the fixed stress of MedFr. The first change reduced all proparoxytones to paroxytones, and the second either reduced the final vowel of paroxytones to schwa, or eliminated it altogether. The loss of proparoxytone penults seems to have occurred over a long period of time. Some loss is attested in vulgar Latin (e.g. TABULA > vulgar Latin <tabla> in the 3rd- or 4th-century *Appendix Probi*; Pope, 1952: §262). Fouché (1952–69: 466–71) dates the loss of most proparoxytones to the 5th century onwards, following the Frankish invasions of northern Gaul. However, some proparoxytonic forms are regularly retained in orthography up until the end of the 12th century:[24]

(37) IMA.GI.NEM > <ima.ge.ne>

(38) VIR.GI.NEM > <vir.ge.ne>

(39) AN.GE.LUM > <an.ge.le>

According to Fouché (1952–69: 472), these forms persist since deletion of the penultimate schwa would produce an impossible consonant cluster (e.g. /dʒn/ or /dʒl/). Eventually, it is not the penult that is deleted, but the final syllable (Pope, 1952: §644). This gives rise to ModFr (and post-12th century MedFr) *image*, *vièrge*, *ange*. However, it is not clear whether the orthographic forms in (37–39) represent true proparoxytones in the 11th and 12th centuries. Fouché observes that when they occur in verse, the two final orthographic syllables count only as one (1952–69: 472). Yet this would leave the orthography anomalous within the MedFr system, which is thought to be broadly phonemic at this time (cf. Pope, 1952: §688–§701). While there could be orthographic influence from church Latin, it is also possible that such proparoxytone forms were still a comparatively recent loss in the 11th century.

23. Dupoux and Peperkamp (2002) argue that the stress rule of southern French varieties, where final schwa is retained, can be extracted by a purely phonological generalization.

24. Dating from quotations in Tobler and Lommatzsch (1925–).

The second change which caused the development of a regular fixed-stress rule was the loss or reduction to schwa of final post-tonic vowels. This had most likely been completed by the time of the earliest vernacular text, the *Strasbourg Oaths* in 842. Forms such as <amur> (from A.MŌ.REM) and <di> (from DĪ.EM) show the loss of final vowels; those such as <cosa> (from CAU.SAM) show that final unstressed /a/ was not lost, while <fradra>, <fradre> (from FRA.TREM show the use of schwa as a support vowel after the <dr> cluster (probably realized /ðr/).[25] Indeed, Fouché (1952–69: 501) notes that the /a/ > /ə/ change may be attested as early as the 6th- to 7th-century *Reichnau Glosses*, since the sound is not consistently written as <a> (Pope, 1952: §260).

From the limited textual evidence available, while it is clear that vowel deletion and reduction was widespread by the time of the earliest texts, it is not clear when it became exceptionless. Fixed word stress emerges as a *result of the completion* of the regular linguistic changes discussed above. Evidence that these changes are in progress at the time of the *Reichnau Glosses*, for example, does not imply that they have run to completion, and exceptions (such as proparoxytones) may still remain. For example, in the *Passion de Clermont*,[26] there is evidence from the scansion that some proparoxytone forms may have been retained:

(40) L'**an.ge.les** Deu de cel des.send (l. 393)

Assuming that the scansion of the line was regular, this line supposes a trisyllabic pronunciation of <angeles>. However, elsewhere a paroxytonic form of the same word is used:

(41) Sus en la pe.ddre l'**an.gel** sist (l. 400)

It is unclear how this variation should be interpreted: it could be the case, for example, that line 393 is simply irregular. This interpretation is perhaps favoured given one further piece of evidence from this text which strongly suggests that a regular word-stress rule has emerged: regular final stress has spread to the pronunciation of Latin (cf. also Pope, 1952: §648).

(42) Il li respondent tuit adun:
"Jesúm querem **Nazarenum**!" (ll. 135–36)

25. Whether <a> represents [a] or schwa [ə] is disputed. Orthographic variation in the spelling of <fradre> most probably shows that the scribe was uncertain how to transcribe the /ə/ support vowel (Ayres-Bennett, 1996: 24). However, Castellani (1969, 1978) argues that the text is from Poitou, and noting that all instances of etymological post-tonic /a/ are written <a>, suggests that they may also have been pronounced as [a].

26. *c.*1000, ed. Avalle (1962). For all texts cited in the present study, a date of composition is given. For early texts this can be very approximate, cf. chapter two, §1.3.1. Where texts are not used in my own corpus, the dating given is that of the *Dictionnaire étymologique de l'ancien français* (Möhren, 2007).

While the diacritic on *Jesúm* may be editorial, the assonance[27] of *adun* and *Nazarenum* suggests non-etymological final stress on the borrowed Latin genitive. A similar stress shift is observed later in the poem on a borrowed Latin verb form:

(43) "Crucifige, **crucifige**!"
crident Pilat trestuit ensems. (ll. 227–28)

Here, the assonance of *crucifige* and *ensems* suggests that both have an oxytonic pronunciation. In the case of the French form *ensems*, this is etymological; in the case of the Latin form *crucifige*, it is not. The oxytonic pronunciation of Latin by French speakers was noted by Palsgrave in the 16th century; moreover, for Peperkamp (2004), the assimilation of non-native vocabulary to a regular stress pattern is taken as evidence for stress deafness. While I remain uncertain at present as to whether 11th-century French speakers showed stress deafness effects, I do consider the oxytonic pronunciation of Latin to be very strong evidence that a regular word-stress rule existed by the early 11th century.

3.1.2 Stress-conditioned allophony

In the development of MedFr from vulgar Latin, there are a number of sound changes which begin life as allophonic variants conditioned by stress. These changes apply to content words in all contexts, and thus provide evidence for the continued existence of word stress.[28]

One such stress-conditioned change is the formation of diphthongs in primary stressed syllables. The four vulgar Latin mid-vowels (/e, ɛ, o, ɔ/) all break and form diphthongs when in syllables with no coda consonant (open syllables). In the same environment, vulgar Latin /a/ becomes /æ/ (written <e>):[29]

(44) PĔ.DEM > vulgar Latin /ˈpɛ.dɛ/ > <pied> /ˈpiɛθ/

27. A simple form of rhyme in which only the stressed vowels are similar, for example *dreit* : *feiz* : *aveir* (/eit/ : /eis/ : /eir/) (*Gormont et Isembart*, ll. 109–11). As final consonants are ignored, Elwert (1965: §111) characterizes assonance as 'homophonie imparfaite'. Robson (1961: 17) suggests that the *Passion de Clermont* shows 'monorimes approximatives', which do in fact take some account of following consonants. It is uncertain how 'similar' the vowels were required to be in assonance, in particular whether or not nasalization was sufficient to distinguish vowels. See Pope (1952: §§428–78) and Rochet (1976) for two different views of the exactness of assonance with respect to vowel nasalization.

28. Some function words develop both stressed and unstressed forms: these are discussed below.

29. The precise phonetic value of this vowel is disputed. I follow Price (1998: 66) in using /æ/; Pope (1952: §233), assumes the value /eː/. All that is important for the present thesis is that it is an <e> which does not rhyme or assonate with other <e>s.

(45) SŎ.ROR > vulgar Latin /ˈsɔ.rɔr/ > <suor> /ˈsuɔr/

(46) TĒ.LAM > vulgar Latin /ˈte.la/ > <teile> /ˈtei.lə/

(47) FLŌ.REM > vulgar Latin /ˈflo.rɛ/ > <flour> /ˈflour/

(48) MAR.EM > vulgar Latin /ˈma.rɛ/ > <mer> /ˈmær/

Chronologically, diphthongization is attested by the time of the *Sequence of Saint Eulalia*, (*c.*880–82).[30] As diphthongization of vulgar Latin /ɛ/ and /ɔ/ (44–45) is attested in other Romance varieties, the process is assumed to have started in vulgar Latin. In French, the process also affects Frankish loan words, so it must still have been active in the 5th and 6th centuries (Pope, 1952: §229; Fouché, 1952–69: 223). However, developments of the vulgar Latin mid-close vowels /e/ and /o/ and the low vowel /a/ are only attested in the *langue d'oïl* region. Frankish influence is usually cited as the trigger for the process, and Fouché (1952–69: 229) dates the developments to the 7th century.

As with the emergence of a regular stress rule, what is crucial from our perspective is not the date at which diphthongs are attested, but the date at which diphthongization ceased to be an active rule. If the rule applied regularly, diphthongs could be analysed as allophones conditioned by word stress: for example, [ei] and [e] could be allophones of a single phoneme /e/. For the mid-open vowels /ɛ/ and /ɔ/, Fouché (1952–69: 221–22) notes that some late Germanic borrowings do not show diphthongization, such as Frankish **breka* > MedFr *breche* /ˈbrɛtʃə/; Frankish **skroda* > MedFr *escroe* /esˈkrɔə/ 'a piece of cloth'. These must have been borrowed sufficiently early to show consonantal changes such as palatalization before /a/ and loss of intervocalic stops, which would suggest that diphthongization of mid-open vowels ceased to be an active rule well before the 11th century. In the case of the mid-close vowels /e/ and /o/, all that can be said is that the rule was no longer active by the 12th century. By this point, central, northern and eastern dialects show a differentiation of [ei] to [oi] — a far less likely allophone of /e/ — while in Western dialects the diphthongs have been re-monophthongized, removing any allophonic variation (Pope, 1952: §230).

From this discussion of diphthongization, we conclude that there is definitive evidence of stress-conditioned allophonic variation until the 7th century, and that it is not attested in the 12th century. It is simply not clear when in the intervening period these stress-conditioned allophones became separate

30. Primary diphthongs of this kind are not attested in the *Strasbourg Oaths* from 842. This may be due to conservative or Latinate orthography, or possibly to dialectal factors, if, as argued by Castellani (1969, 1978), the text is Poitevin.

phonemes.[31]

The effect of stress on the development of a number of function words varies according to context. Under full tonic stress, the vowels in the personal pronouns MĒ TĒ and SĒ diphthongize to produce *mei*, *tei* and *sei* (from the 12th century, *moi*, *toi* and *soi*).[32] These are known as 'strong forms' (Moignet, 1965; de Kok, 1985). However, weak forms also exist, reflecting historically unstressed developments. These are written <me>, <te> and <se> and pronounced with final /ə/. Weak forms may only be found adjacent to the finite verb, either pre- or post-verbally (Foulet, 1930: §151; de Kok, 1985: 48). Phonologically, weak forms evolved as clitics (cf. de Kok, 1985: 152–82), and their restricted syntactic distribution in MedFr suggests that their clitic status was not simply phonological.[33] Unstressed clitic forms may be found in both word- and group-stress languages, and so this evidence that function words could be unstressed in certain positions in the clause does not indicate the presence of a group-stress system. I leave the case of the use of strong forms in post-verbal position to a fuller analysis in chapter four; the use of strong forms in pre-verbal position is discussed in section 3.2.

Evidence for stress-conditioned allophony may also be found in patterns of vowel reduction to schwa. From the 11th century, pretonic /e, ɛ, iɛ, æ/ in open syllables are reduced to [ə]. Pope (1952: §§234–35) describes this as a change mainly affecting initial syllables:

(49) NĔ.PŌ.TEM > vulgar Latin */nɛ.ˈpo.te/ > /**nə**.ˈveu/

(50) DĔ.BĒ.RE > vulgar Latin */dɛ.ˈβe.re/ > /**də**.ˈvoir/

(adapted from Pope, 1952: §234)

In both cases, initial /e/, descended from vulgar Latin /e/ or /ɛ/, has been

31. The subsequent development of diphthongs does not provide such clear evidence for the existence of word stress. Pope claims that the differentiation of /ei/ to /oi/ and /ou/ to /eu/ in the 11th century occurred 'while the tonic stress was still strong' (1952: §226). Yet it is unclear why this kind of differentiation should be incompatible with group stress. The monophthongization of diphthongs in the 12th and 13th centuries (cf. Pope, 1952: §§503–57) may be more plausibly attributed to 'the return to a more level intonation' (1952: §508), i.e., a weakening of primary stress. However, it is less certain that monophthongization is incompatible with a word-stress system.

32. Similar stress conditioning also affected the development of third person pronouns; however, the development of the various morphological forms of ILLE into pronouns and determiners is far more complex and more widely debated: cf. Zink (1992: 91–92), Pope (1952: §§833–43) for brief overviews.

33. Matthews (2007) defines a clitic as 'any grammatical unit that is not straightforwardly either an affix or a word on its own.' Cliticization may have both prosodic and syntactic effects; however, following Klavans (1985), I assume that syntactic and prosodic cliticization are independent processes. Where the term 'clitic' is used in the present thesis, it is to be understood phonologically: i.e. a non-affixal grammatical unit not parsed as an independent prosodic word.

reduced to schwa. Unfortunately, as orthography never reflects the change, evidence is scant: Pope notes only a transliteration of the French word *reine* in Cyrillic from 1063, the Cyrillic alphabet having a means to distinguish schwa from other vowels which is absent in the Roman system (1952: §235).

Using Hebrew glosses of 13th-century French, Morin (1991) finds some evidence that a vowel reduction rule may affect the phonemes /e, ɛ, iɛ, æ/ in all positions in the word. The glosses suggest that alternate instances of these vowels are reduced to schwa. Morin considers three forms of the verb *reveler*:

(51) (a) *[rəˈvɛlə] (third person plural present indicative)
(b) <rēvəler> (infinitive)
(c) <rəvel[ə]reç> (written <rəvelareç>) (second person plural future)
(Morin, 1991: 59)

Note here that all examples contain full vowels alternating with vowels reduced to schwa. However, the rule is not consistently observed in Morin's data. He suggests that while vowel reduction was at one stage a regular rule, this was no longer the case in the 13th century. Its regularity was perhaps disrupted by the monophthongization of /ai/ and /ei/ diphthongs, producing an /ɛː/ vowel which does not seem to have been subject to vowel reduction (1991: 71). [34]

3.1.3 Summary

To summarize, in section 3.1, I have examined two phonological changes in the development of French which help to establish a *terminus a quo* for the development of group stress. A fixed word-stress rule, a necessary pre-condition for the emergence of group stress, was found to exist in the early 11th century. Stress-conditioned allophony at the word level is lost by the 12th century at the latest. In short, we have seen no evidence which either proves or falsifies the hypothesis that group stress existed in the 12th century, and some evidence which suggests that it could have existed as early as the 11th century.

3.2 Evidence from morphosyntactic change? The 'strong initial position'

It is often claimed that the initial position in the clause in French of the 12th and 13th centuries was prosodically strong. Marchello-Nizia (1995) traces

34. Morin (1991: 58–61) also suggests that the rule was never conditioned purely by the position of primary stress, but was also affected by the position of vowels which could not be reduced to schwa.

the claim back to Thurneysen (1892), while Adams (1989: 14) states that it is 'a cliché of the handbooks'. Both scholars adopt the assumption in their own work. For Kukenheim (1971) too, the loss of phenomena which show initial as opposed to final stress is taken as evidence for the prosodic change from 'ascending' to 'descending' rhythm.

In this section, I examine two phenomena which are commonly argued to be affected by the 'strong initial position': the verb-second syntax of MedFr and the position and form of pre-verbal object pronouns. In the framework of the present study, I argue that it is problematic to see these phenomena as inherently 'stress conditioned' in the same way as the segmental phonological processes discussed above. Consequently, I suggest that there is still work to be done to establish whether or not morpho-syntactic changes in these areas are related to the emergence of group stress, a problem to which I return in chapter four.

3.2.1 Verb-second effects

In MedFr of the 12$^{\text{th}}$ and 13$^{\text{th}}$ centuries, the finite verb of the main clause is frequently the second constituent, a phenomenon known as verb-second (V2).[35] The initial constituent of the sentence may be any other element. (Note that in MedFr, object pronouns and the pre-verbal negative *ne* are considered clitics on the finite verb and do not count as separate constituents. Pre-verbal and verbal constituents are delimited by brackets.)

(52) [**Ele**] [vos a] atendu lonc tens
(subject pronoun; *La Queste del Saint Graal*,[36] p. 180, l. 9)

(53) [**Li deables**] [m'a] mostree la douçor et le miel
(nominal subject; *ibid.* p. 65, l. 3)

(54) [**Einsi**] [parla] la voiz au roi
(adverb; *ibid.* p. 86, l. 4)

(55) Et [**de ces trois choses**] [vos diré] je bien
(prepositional phrase; *ibid.* p. 37, l. 29)

(56) [**Ceste costume**] [ai] je toz jors tenue
(nominal direct object; *ibid.* p. 5, l. 7)

35. From the initial analysis of Thurneysen (1892), this observation has been confirmed and elaborated by many other scholars, e.g. Herman (1990), Skårup (1975), Marchello-Nizia (1995); in a typological framework Harris (1978), Buridant (1987, 1992), Combettes (1988); in a generative framework Adams (1987a,b, 1989), Vance (1988, 1995b, 1997), Roberts (1993), Lemieux and Dupuis (1995), Labelle and Hirschbühler (2005), Labelle (2007).

36. *c.*1220, ed. Pauphilet (1923).

(57) [**Bel**] [sont] il voirement

(attributive adjective; *ibid.* p. 156, l. 22)

(58) Perceval, [**traïe**] [m'avez!]

(past participle; *ibid.* p. 110, l. 18)

(59) [**a estre principaus compains des compaignons de la Queste**] [eussiez] vos failli

(non-finite clause; *ibid.* p. 80, l. 6)

(60) Et [**por ce que tu n'ailles sels**,] [voil] ge que tu meines o toi Perceval et Boorz

(subordinate clause; *ibid.* p. 271, l. 13)

(examples from Vance, 1997: 39–46)

The examples above show that the vast majority of clausal constituents can fill the initial position. Indeed, the only ones that cannot are conjunctions (e.g. *et* in 55, 60; *mais*, *ne*, *car*),[37] clitics on the finite verb (e.g. *vos* in 52) and the pre-verbal negative *ne*.

Thurneysen (1892) proposed that the word order of MedFr could be explained by rhythmic factors. Each clause has a strong initial stress, and this stress could not be realized on a weak element; and for Thurneysen (1892), the finite verb was prosodically weak. The verb was therefore placed in second position, enclitic upon a preceding strong element (Thurneysen, 1892: 300). Yet there is a simple problem of cause and effect here: does an obligatory initial stress cause a strong element to appear in initial position, or does a strong element in initial position cause an obligatory initial stress? Variations on both positions have been defended for more than a century: Richter (1903: 45–51), for example, argues that the initial position is characterized by an element which links the clause to the preceding context, similar to what would today be called a 'topic', and that accentuation is a consequence of this. Herman (1990) adopts an intermediate position. On the one hand, prosodic structure is fixed, and this favours a particular syntactic structure:

> L'existence d'un accent syntaxique initial *influe sur le choix du premier terme de la proposition.*
>
> (Herman, 1990: 285; my emphasis)

37. Lemieux and Dupuis (1995: 95–100) argue that initial *et* can count for V2 in the 12$^{\text{th}}$ century. I follow Vance (1997: 74–75), who argues that *et*–V constructions have a different underlying structure, noting in particular the impossibility of verb–subject pronoun inversion in such cases.

On the other hand, this initial stress cannot be purely rhythmic, as Thurneysen suggests:

> La structure rythmique de la proposition ne peut être considérée en elle-même [...] L'accent initial, tout en étant un phénomène phonétique, est étroitement lié à des facteurs sémantiques et syntaxiques.
>
> (Herman, 1990: 285)

The most recent defender of the 'rhythmic V2' hypothesis is Adams (1987b, 1989). Adams argues that there is a prosodic requirement in languages with 'heavy stress' for the initial constituent to be stressed:

> **Eurhythmic Principle**
> In languages with heavy stress, place stress on the initial constituent.
>
> (Adams, 1989: 9)

Adams assumes that for syntactic reasons the finite verb must occur towards the beginning of the clause (cf. chapter four, §3.1.1). However, if this fixed initial stress is realized on the finite verb, pragmatic effects ensue. Thus verb-initial clauses are associated with marked clause types placing special emphasis on the verb (imperatives, interrogatives, 'lively' narrative).[38] Consequently, stress on the verb is avoided by placing another constituent in initial position. Adams argues that a pragmatic restriction (i.e. the initial constituent is pragmatically marked, and pragmatic marking entails stress) cannot be correct. Firstly, she suggests that in both German and MedFr, 'except where the object is fronted, the initial constituent receives no necessary marked or emphatic interpretation' (Adams, 1989: 4). As evidence for this, Adams highlights the existence of pre-verbal expletive pronouns in German, which cannot have a pragmatic meaning (1989: 12). Secondly, if only pre-verbal objects are pragmatically marked in initial position, we may expect that pre-verbal subjects can be unstressed. Adams cites evidence from Frisian, where both a full (*hy*) and a clitic form (*er*) of the subject pronoun is available, to show that in fact only the full form is permitted in clause-initial position:

(61) (a) Pyt sei dat hy/er my sjoen hie.
Pyt said that he me seen has

(b) Hy/*er hie my sjoen.
He has me seen

(Adams, 1987b: 143, from de Haan and Weerman, 1986)

The rhythmic verb-second hypothesis as defended by Adams is flawed in a number of important respects. Firstly, the eurhythmic principle is an entirely

38. See, for instance, Roberts (1993: 56–58).

unproven stipulation, based (as far as I am aware) on the observation that Germanic has 'heavy stress' [39] and that stress generally falls on the initial constituent. Two arbitrary typological predictions are made: firstly, 'heavy' stress and a *lack* of stress on the initial constituent is impossible, and secondly, 'non-heavy' stress and regular stress on the initial constituent is impossible. Based on the second of these assumptions, Adams (1989: 27) argues that the loss of verb-second is caused by the loss of heavy stress in MedFr. In short, while the eurhythmic principle appears to explain the facts of German and MedFr, there seems to be no principled linguistic reasoning underlying it. Secondly, the notion that a prosodic rule may be sensitive to syntactic constituency is troubling. While prosodic constituents may be derived from the syntax (cf. §1.1), they are not necessarily co-extensive with syntactic constituents. Yet MedFr V2 is clearly sensitive to syntactic constituency, as shown by the variety of elements which may precede the finite verb. Particularly interesting is example (59) above, where the finite verb is preceded by a infinitival clause *a estre principaus compains des compaignons de la Queste*. While the whole clause forms a single syntactic constituent at the level of the matrix verb, it may also be subdivided into constituents. The first of these is the infinitival copula *a estre*. It seems unusual that a prosodic rule which avoids stress on all finite verbs should not reject a word order where an infinitival copula appears in clause-initial position.

In summary, I reject the argument that verb-second is *caused* by a prosodic constraint requiring a strong initial stress. At present, therefore, we have no reason to associate loss of V2 with loss of group stress. However, the loss of group stress may have triggered a chain of events resulting in a linguistic system where the V2 constraint was more difficult to acquire. This hypothesis is examined in chapter four (§3).

39. The term 'heavy stress' is problematic. Adams seems to be drawing on scholars such as Pope, who refers to a 'gradual lessening of heavy tonic stress' in the development of French (1952: §170). Pope characterizes stress purely in articulatory terms, with expiratory stress being 'produced by variations in the force with which air is expelled from the lungs' (1952: §118). Unfortunately, I am not aware of any means of measuring the physical articulation of stress to determine whether it is 'heavy' or 'light'. Pope associates heavy stress with the lengthening and diphthongization of stressed vowels, and reduction or deletion of unstressed vowels (1952: §121). As Dauer (1983: 59) points out, these properties are common in languages which are argued to be *stress timed*. Indeed, approaches for measuring linguistic rhythm focus on relative differences in syllable length (Ramus et al., 1999; Grabe and Low, 2002), with languages that show greater variation in syllable length deemed to be more 'stress timed'. Thus, for the 'heavy stress' suggested by Adams, I understand 'strongly stress timed'.

3.2.2 The form of pre-verbal non-subject pronouns

In the French pronominal system of the 12th century, a number of forms were represented by several allomorphs.[40] All pronouns had a 'weak' form:[41]

MedFr weak pronouns
me (1SG), *te* (2SG), *nos* (1PL), *vos* (2PL, *se* (REFL)
le (3SG.MASC.DO), *la* (3SG.FEM.DO), *les* (3PL.DO)
li (3SG.IO), *lor* (3PL.IO)
(DO = direct object, IO = indirect object)

As discussed above, *me*, *te* and *se* had strong allomorphs (*moi*, *toi* and *soi*), which did develop under primary stress. To *les* corresponded two strong allomorphs: *eus* (masc.) and *eles* (fem.). Other strong forms existed; however, these were not exact allomorphs of any weak form (e.g. *lui*, which was initially an allomorph of both *le* and *li* (masc.), but not *li* (fem.)). However, there was also a third series of pronouns, often called 'enclitics' (e.g. Pope, 1952: §838), which I refer to as 'consonantal forms':

MedFr consonantal pronouns
m (1SG), *t* (2SG), *s* (REFL)
l (3SG.DO), *s* (3PL.DO) (de Kok, 1985: 66)

Until the 13th century, weak forms do not appear in clause-initial position. This phenomenon, observed in all medieval Romance varieties (cf. Benincà, 1995, 2006), is known as the Tobler-Mussafia law (Tobler, 1875; Mussafia, 1886). Non-subject pronouns are thus post-verbal when the verb is the first constituent in the clause.

(62) Ot **le** Guillaumes, s'est vers terre clinez
(*Le Couronnement de Louis*,[42] l. 1403; Foulet, 1930: §162)

Prosodic explanations for the second position of the clitic were widely accepted until the second half of the 20th century: Foulet, for example, explains that the pronoun is post-verbal 'pour éviter de faire tomber sur [lui] un accent trop marqué' (1930: §162) (cf. de Kok, 1985: 60–62). A second consequence of the strong initial position is that where a non-subject form does appear in pre-verbal position, a strong allomorph must be used:

40. Cf. de Kok (1985: 17–27), Moignet (1965: 49–60) or Buridant (2000: §326) for a more detailed presentation.
41. I.e. a form which did not develop under tonic stress (Pope, 1952: §§833–37). Despite the fact that they may also appear as subject pronouns and after a preposition, *nos* and *vos* do not show the expected stressed development, **neus* and **veus* (Moignet, 1965: 54).
42. Second third of the 12th century, ed. Langlois (1925).

(63) quant **moi** plera

(*La Mort le roi Artu*,[43] p. 66 l. 32; de Kok, 1985: 54)

Moignet (1965: 65) suggests that this syntax is most common with impersonal verbs; from the examples he gives, the non-subject pronoun seems normally to represent the experiencer. These verbs form a plausible syntactic exception both to the Tobler-Mussafia law and to the verb-second constraint, since here, as Moignet (1965: 66) points out, the non-subject pronoun is semantically the 'thème' (or topic) of the proposition, much as is the subject of a non-impersonal verb. However, the use of a strong form of the non-subject pronoun here may suggest a rhythmic constraint imposed by a strong initial position.

Yet there is a fundamental flaw in the rhythmic approach to the Tobler-Mussafia law, first signalled by Ramsden (1963: 124). By the 12th century, post-verbal pronouns too have come to be realized by a strong form. Moignet (1965) apparently sees no contradiction when he writes:

> On sait que des *raisons de rythme* provoquent, dans certains cas précis, la postposition du pronom personnel régime. Le pronom personnel proprement dit *apparaît alors sous la forme prédicative moi*, *toi*, *soi*.
>
> (Moignet, 1965: 66–67; my emphasis)

It seems implausible that the requirement for a phrase-initial stressed form should cause a weak pronoun to occur in post-verbal position, only for it then to be realized as a strong form. Ramsden, noting the exceptional cases of strong pre-verbal pronouns such as *moi plera* above, argues that 'it is not the presence of a tonic [i.e. strong] form in exceptional and emphatic position that should interest us, but its appearance in post-verbal position where traditionally the unstressed pronoun had been used' (1963: 124). Thus the replacement of a weak with a strong allomorph in *post-verbal* position may well be connected to the emergence of group stress, and we will return to the issue in chapter four (§2).

In short, the idea that the Tobler-Mussafia law is caused by a prosodic requirement stipulating that weak forms of pronouns cannot occur in initial position must be rejected.[44] The loss of Tobler-Mussafia effects in French from

43. First quarter of the 13th century, ed. Frappier (1954).

44. Whether an alternative 'explanation' exists is a more difficult question. Generative approaches (e.g. Benincà, 1995, 2006; Cardinaletti and Roberts, 1991; Labelle and Hirschbühler, 2005) suggest that in verb-first constructions, the verb moves higher in the clause structure to be focalized for discourse purposes. It is therefore a property of the verb, rather than of the pronoun, that leads to post-verbal position of non-subject pronouns in these cases.

the late 12th century, at which point weak pronouns do begin to be attested in clause-initial position (de Kok, 1985: 90–91), cannot be automatically associated with the emergence of group stress, and certainly cannot be used to date this prosodic change. More generally, I agree with Benincà, who argues that 'it has never been reported, nor can be expected on theoretical grounds, that phonological constraints feed syntactic movements' (1995: 331).

The alternation of weak and consonantal allomorphs, on the other hand, clearly derives from a phonological rule. Where a non-subject pronoun follows a vowel, it is realized as a consonantal form.[45] The examples below show the consonantal form *l* following a number of constituents:

(64) **nel** reconurent li dui sergant sum pedre
(negative; *Vie de Saint Alexis*,[46] l. 117)

(65) plainums ensemble le doel de nostre ami:
tu tun seinur, **jol** f[e]rai pur mun fils (subject pronoun; *ibid.*, l. 155)

(66) se lui'n remaint, **sil** rent as poverins (adverb; *ibid.*, l. 100)

(67) n'est hom **kil** veit (complementizer; *Chanson de Roland*,[47] l. 530)

De Kok (1985: 66–67) outlines a chronology for the disappearance of consonantal forms following a vowel, and their replacement by a weak form. In the 11th century, all consonantal forms are attested after any word ending in a vowel, although Pope (1952: §602) argues that they were rare after words of more than one syllable (e.g. *en* ***terrel*** *metent*; *Vie de Saint Alexis*, l. 588). In the 12th century, only the forms *l* and *s* are found, and then only after the monosyllables *ne*, *jo/je*, *si*, *se*, *que*, *qui*, *ja*, *ou*, *issi*, *ço* and *tu* (de Kok, 1985: 66). In the 13th century, consonantal forms are only found after *ne*, *jo/je* and *si*. For Pope, the disappearance of consonantal forms after a vowel may have a stress-based explanation:

> With the lessening of the intensity of tonic stress the repugnance to juxtaposition of unstressed words diminished and the practice of enclisis [...] gradually died out.
>
> (Pope, 1952: §602)

It is undeniable that these forms must have developed in an unstressed position.[48] Whether they were still derived by a productive vowel deletion rule

45. Consonantal pronouns are also found where a non-subject pronoun *precedes* a vowel; as this process of elision is still found in ModFr, it will not be further discussed.

46. End 11th century, ed. Storey (1968).

47. *c.*1100, ed. Bédier (1928).

48. Whether as enclitics, proclitics or intertonic syllables has been debated for over a century: see de Kok (1985: 152–82).

in the 11th and 12th centuries is another matter. This is the position assumed by Horne (1990a), who argues that the enclitic pronoun, its host, and the verb all form part of the same prosodic constituent, the clitic group of Nespor and Vogel (1986):

(68) [jo te vi]$_{\text{ClGrp}}$ (analysis of Horne, 1990a)

This prosodic constituent 'retained the initial stress characteristic of Gallo-Roman words' (1990a: 7), i.e. the secondary stress realized on initial syllables in vulgar Latin. Within the clitic group, a productive vowel syncope rule took place, deleting intertonic vowels: [49]

(69) [jò te ví]$_{\text{ClGrp}}$ > /ˌdʒot ˈvi/

Most interesting from Horne's analysis is the idea that a constituent higher on the prosodic hierarchy than the PWd could have a fixed stress pattern, a position which I defended for ModFr in section 1. However, it is unclear how or why the clitic group should have developed a fixed stress pattern. [50]

Dufresne (1993, 1995) and Dufresne and Dupuis (1994) assume that weak pronouns are attached to the preceding word at the level of the PhPh. Where the rhyme of the preceding syllable is empty, the initial consonant of the clitic moves to fill it. Dufresne (1995: 101) supposes that a vowel-deletion rule then eliminates the schwa vowel of the pronoun. [51] Such a rule, the authors argue, is dependent on strong word stress. Yet if there is a vowel deletion rule, it is not clear why this should apply to the final vowel of the clitic *te*, but not to word-final schwa. In other words, if *jo+te* is realized *jot* due to a preceding stressed vowel, why is *rote* 'troupe' not realized **rot*?

To summarize, while consonantal forms of non-subject pronouns certainly derive from a historical process of vowel deletion, analyses which suppose the vowel deletion rule to remain productive in the 11th and 12th centuries are problematic. In particular, we note that de Kok's (1985) chronology for the disappearance of consonantal forms suggests that host plus pronoun combinations are lexicalized by this point, being restricted to a small number of host words and an even smaller number of consonantal pronouns. Moreover, if a vowel deletion rule is not active at this time, we cannot then link the disappearance of enclisis to a lessening of the intensity of stress. Indeed, else-

49. Cf. Pope (1952: §§234–35) for the operation of such a rule at the word level in the Gallo-Roman period.

50. The treatment of the subject pronoun as a clitic in MedFr is also controversial, cf. Jacobs (1993: 153, 160–61).

51. 'Les voyelles faibles sont réduites ou effacées lorsqu'elles sont en présence de voyelles accentuées' (Dufresne, 1995: 101–2).

where Pope herself proposes a far more plausible reason for the loss of these consonantal pronouns:

> The phonetic changes of the 12th century (the vocalization of /l/ and the effacement of prae-consonantal /z/) often obscured the significance of these forms.
>
> (Pope, 1952: §838)

We conclude that phonetic rather than stress-based factors account for the loss of consonantal pronouns in the medieval period.

3.2.3 Summary

An examination of the morpho-syntactic developments associated with the alleged strong clause-initial position in French of the 12th and 13th centuries highlights a number of very questionable assumptions that have been made in previous studies. Firstly, I have rejected the claim that there is an underlying 'rhythmic template' for the MedFr clause. If it is the case that the initial element of the clause is usually stressed, it seems more plausible that this is the consequence of syntactic movement rules, perhaps with pragmatic triggers. Secondly, it follows trivially from this claim that the loss of V2 word order and changes in the forms of non-subject pronouns cannot be attributed to the loss of word stress. This is not to say that morpho-syntactic changes *cannot* occur as the result of prosodic change, merely the link assumed by previous studies is not correct. Since a more complex argument is required if morpho-syntactic data is to be used, we may draw no conclusions from it at present, and postpone a fuller analysis to chapter four.

4 Evidence from versification

In the previous section, I examined indirect evidence from segmental phonology and morpho-syntax to attempt to gain insight into the emergence of group stress in French. Based on segmental phonological evidence, I suggested a *terminus a quo* of the early 11th century for the emergence of group stress, but there exist few segmental changes that can be used as evidence for the nature of stress beyond this point. While chronologically later, morpho-syntactic developments were shown to be a problematic source of evidence as to the nature of stress.

In this section, I investigate a different approach to the problem of stress, exemplified by studies of the rhythm of verse. Rather than studying related linguistic phenomena, studies of verse reconstruct the position of stress to assess whether there is evidence that poets are using stress to structure the verse. In particular, I examine studies which suggest that early MedFr versification was syllabo-tonic (i.e. both the syllable count and the position of stressed syllables were metrically relevant). From the point of view of the present study, I will argue that the existence of syllabo-tonic versification would provide crucial evidence about the perception of stress in MedFr. In section 2, I argued that group stress (a production rule) and stress deafness (a perceptual / processing effect) were two sides of the same coin. Here, I suggest that stress deafness and syllabo-tonic versification are incompatible, and thus evidence from verse constitutes the most promising starting point for an enquiry into the emergence of group stress.

4.1 Syllabic meter

MedFr versification is based on syllable count (Elwert, 1965: §29). Lines are written so as to contain an equal number of syllables,[52] with the most common line lengths being eight, ten and twelve syllables (Elwert, 1965: §§160–62). The final syllable of the line bears an obligatory stress, and rhymes or assonates with the final syllable of neighbouring lines. For reference, I reproduce a typical eight-syllable (octosyllabic) extract below:

(70) Na.gai.res, che.vau.chant, pen.soy.e
Com ho.me tri.ste‿et do.lo.reux,
Au dueil ou il fault que je soy.e
Le plus do.lent des a.mou.reux[.]

(Alain Chartier, *La Belle Dame sans mercy*,[53] ll. 1–4)

It is standard in MedFr versification for line-final schwa not to be counted: thus in the first and third lines of (70), the final schwa in *pensoye* and *soye* is the ninth syllable of the line. Consequently, the eighth syllable of the line is always a primary stressed syllable.

With the possible exception of the very earliest texts, this is the only clear metrical constraint on the position of stress in the octosyllabic line. Lote takes the view that all syllables not at the rhyme are unstressed:

52. The later Anglo-Norman tradition is a notable exception; see Johnston (1980, 1983).
53. 1425, ed. Laidlaw (1974).

> A l'intérieur des hémistiches, toutes les syllabes sont primitivement égales entre elles et atones, par opposition à la césure[54] et la rime qui sont toniques.
>
> (Lote, 1949–96: I, 306)

Lote's evidence is drawn from the musical settings of medieval texts, which in the 13th century show a variety of strict 'modal rhythms' which can be applied to any text. If a text can be set to a variety of rhythmic patterns, argues Lote, then it cannot have any fixed rhythmic properties of its own (Lote, 1949–96: I, 306–47). However, it seems unlikely that octosyllabic verse contains a sequence of seven unstressed syllables, as analyses of ModFr suggest that this would constitute an abnormally long stress group.[55] Elwert suggests instead that 'le vers peut avoir jusqu'à trois syllabes accentuées, dont l'intensité et le point d'application sont variables' (1965: §162). Nevertheless, metrically, these two points of view are identical: there is no metrical restriction conditioning the placement of stressed syllables.

In the remaining two sections, two counter-claims to this position will be examined. Firstly, I will discuss evidence that French versification was syllabo-tonic in the 11th century and early 12th century, with a meter which required regular alternation of stressed and unstressed syllables. Secondly, I will consider the nature of a metrical constraint present in ten- and twelve-syllable verse, the cæsura, and discuss evidence that the earliest octosyllabic texts may also have shown such a constraint.

4.2 Syllabo-tonic meter

There are two principal counter claims to this view which are of interest to us. The first is represented by studies which claim that there is some organization of stressed syllables in French versification (e.g. Klausenburger, 1970; Le Mée, 1978; Pensom, 1982, 2000, 2009; Noyer, 2002). In other words, early French verse shows a syllabo-tonic meter. Noyer (2002) considers a substantial corpus of verse texts, which contains 42 distinct works from the earliest octosyllabic texts (*c.*1000) to 1180. A clear diachronic trend in the rhythm of verse is shown. The oldest texts (from the 11th century and early 12th century)

54. I discuss the cæsura below; however, there is no cæsura in the majority of octosyllabic texts (Elwert, 1965: §162).

55. For example, in the analysis of Delais-Roussarie (1995), a constraint MAX limits stress groups to a maximum of six syllables. Post (2000: 91) reports that stress groups of three or four syllables are more usual.

are strongly iambic, while later 12th-century texts show much weaker rhythmic organization. However, all texts contain a greater than chance frequency of iambic lines. [56]

If it is the case that texts from the 11th century and early 12th century show a form of syllabo-tonic versification, this is very significant for the emergence of group stress and corresponding stress deafness effects. For syllabo-tonic versification to be effective, the distinction between stressed and unstressed syllables must be phonologically relevant to speakers of the language. The link between the phonological features of a language and the system of versification used is made by the metricist M. L. Gasparov in the introduction to his *History of European Versification*:

> Every system of versification relies on the phonological resources of a particular language. Pure syllabic and pure tonic versification are possible, it would seem, in any language, if 'tonic' is understood as counting by the number of phonetic words.
>
> (Gasparov, 1996: 4)

By Gasparov's own definition, a group stress system such as that of ModFr would seem to be the exception to the rule that tonic (or syllabo-tonic) versification is possible in any language. [57] We have seen evidence that stress is not phonologized by ModFr speakers, and is therefore not one of the 'phonological resources' of the language. Moreover, 'phonetic words' (in the sense of the minimal stress-bearing unit) do not exist in a group stress system: stress is realized as a property of a higher-level prosodic constituent. Therefore, we do not predict syllabo-tonic versification to be attested in a language with a group-stress system. Consequently, if we accept Noyer's argument that 'the Iambic Pattern, although violable, must still be postulated as part of the metrical system of Old French octosyllabic verse' (2002: 10), we have a strong indication that word-level stress was still present in MedFr until at least the middle of the 12th century.

While the data presented is intriguing, Noyer's (2002) methodology is based on the premise that metrical constraints developed in the study of the English

56. Chance frequency is established with reference to prose, cf. chapter three, §5.1.

57. I agree with Gasparov (1996: 141–42) that it is quite possible to compose iambic verse in ModFr. However, I disagree with the conclusion he draws from this: that because it is possible to write syllabo-tonic verse, the 'resistance of the language' (1996: 141) cannot play an important role in explaining the absence of syllabo-tonic verse forms in ModFr. The reason French speakers are 'not accustomed to paying attention to stress as a unit for measuring verse' (1996: 142) is likely to be connected with stress deafness at least as much as with the syllabic traditions of Romance versification.

iambic tradition had some relevance in MedFr. Unfortunately, Noyer's own data show this premise to be false. Using the framework of Generative Metrics (Halle and Keyser, 1971; Kiparsky, 1977), the investigation tests the conformity of lines to the Iambic Pattern:

> **Iambic Pattern**
> W S W S W S W S (X) (Noyer, 2002: 128)

Two correspondance rules are used:

Polysyllable Rule The main stress of a stressable polysyllabic word, unless verse- or phrase- initial, may occupy only a S[trong] position.

Monosyllable Rule A stressable monosyllabic word, occurring between two stressless syllables [...] within a line of verse, and not preceded or followed by a major phrase boundary, may occupy only a S position.
(Noyer, 2002: 132)

These correspondance rules are devised through the study of English iambic verse by Kiparsky (1977) and Halle and Keyser (1971) to define the minimum requirements for a metrical iambic line. Consider the example of English iambic pentameter below:

(71) And last, **poin**ting to Corinth, ask'd her sweet
(John Keats, *Lamia* I, l. 342; Noyer, 2002: 125)

It is not the case that a stressed syllable always occurs in a strong position, but in this line, the only exception is the initial syllable of *pointing*, which occurs at the beginning of a clause. Such clause- or line-initial foot inversion in English verse is relatively common, and therefore such examples are explicitly excluded from infringing the correspondance rules (cf. Hanson and Kiparsky, 1996: 293). Thus, (71) is considered metrically iambic, despite a minor rhythmic deviation from the pattern. By allowing for rhythmic deviations, these correspondance rules were intended to be *inviolable* constraints on metricality in iambic verse, allowing sufficient variation in rhythm to accommodate all lines of verse in the tradition. While not all poets in the English tradition write verse which conforms perfectly to the correspondance rules (cf. Noyer, 2002: 126–27), particular exceptions adopted by individual poets can be identified and formalized (see, for example, Duffell, 2008: 137–60; Fabb and Halle, 2005: 45–48). However, when these correspondance rules are applied to MedFr texts, violations are commonplace. Only the four earliest octosyllabic texts show fairly strong conformity to the Iambic Pattern: in the *Vie de Saint Léger* (*c.*1000), 85.0% of lines conform; in the *Passion de Clermont* (*c.*1000), 82.6%

of lines; in *Gormont et Isembart* (first half of the 12$^{\text{th}}$ century), 82.1% of lines; in the (very short) octosyllabic fragment of the *Roman d'Alexandre* (*c.*1100), 92.0% of lines (Noyer, 2002: 136). Few other texts show over 75% iambic lines; for many it is fewer than 60%. Noyer's figures may seem initially to support the hypothesis that an iambic syllabo-tonic meter is used in MedFr. Yet there is a critical flaw here. The constraints used by Noyer are already permissive: they do not apply to every stressed syllable, and allow for exceptions at the beginning of syntactic phrases (cf. example 71). Fixed exceptions are included so that the constraints provide an *absolute* definition of metricality in the English iambic tradition. In other words, rather than demonstrate that the Iambic Pattern is relevant to MedFr, Noyer's study demonstrates that a substantial proportion of lines (in many cases, over 40%) are *unmetrical* if the versification of the octosyllable is assumed to be iambic. Noyer (2002: 159) acknowledges this shortcoming, but its implications are fundamental: in effect, the analysis is based on the clearly false premise that English-style versification constraints are relevant to MedFr.

4.3 Cæsura and the octosyllable

Nevertheless, the data provided by Noyer suggests that some rhythmic organization is present in early octosyllabic texts. Guthrie (1987) argues that the primary rhythmic feature of octosyllabic texts is the presence of a cæsura. The cæsura is unquestionably a feature of ten- and twelve-syllable verse in the medieval period. Elwert offers the following definition:

> La césure consiste à faire suivre un nombre déterminé de syllabes, dont la dernière doit être accentuée, par une coupe syntaxique, à laquelle correspond une courte pause.
>
> (Elwert, 1965: §82)

The following example illustrates the cæsura in two lines of ten-syllable verse:

(72) dist O.li.ver: // "pai.en unt grant es.forz;
de noz fran.ceis // m'i semb.let a.veir mult poi!"

(*Chanson de Roland*,[58] ll. 547–48)

In the framework of the present thesis, one of the first things we notice about Elwert's definition of the cæsura is the association of stress, syntactic break, and pause, which recalls the properties of the edges of the ModFr stress group. While the cæsura may be a 'coupe syntaxique', its association with a prosodic

58. *c.*1100, ed. Moignet (1969).

feature (a pause) suggests that it may be better analysed as a prosodic rather than a syntactic constituent.[59] Nespor and Vogel (1986: 280–83) suggest that the PhPh plays a role in determining the position of primary metrical stress in medieval Italian *endecasillabi* (equivalent to French ten-syllable verse). In the Italian *endecasillabo*, the primary stress falls most often on either the fourth or the sixth syllable (1986: 281). In French, the position of the cæsura is more regular, generally falling after the fourth syllable in ten-syllable verse, and after the sixth syllable in twelve-syllable verse (Elwert, 1965: §85). From these observations, I suggest that the cæsura is best defined as a metrical constraint on the position of a PhPh boundary within the line of verse. The types of cæsura found in ten- and twelve-syllable verse will be discussed in chapter two (§2.2).

Elwert (1965: §162) suggests that only the earliest octosyllabic texts contain a cæsura, dividing the line into two sequences of four syllables. A cæsural analysis of the medieval octosyllable is developed by Guthrie (1987). Like Noyer (2002), Guthrie identifies a rhythmic change in the octosyllable, but he ascribes this to changes in cæsural practice rather than to a system of syllabo-tonic versification. Guthrie considers there to be two principal types of fourth-syllable cæsura. Group I corresponds to the definition of cæsura suggested by Elwert: here, there are 'clear phrase boundaries' after the fourth syllable:

(73) au de.par.tir // dou bel es.té
(Guillaume de Machaut, *Jugement dou Roy de Navarre*,[60] l. 1)

(74) pour li ser.vir // et hon.nou.rer (*ibid.*, l. 11)

The earliest text in Guthrie's corpus, *Gormont et Isembart*, contains 79% of Group I lines. Guthrie goes on to argue, however, that a second type of fourth-position cæsura must also be considered, one in which there is a word boundary but no corresponding PhPh boundary:[61]

(75) et met.tent a // des.truc.ti.on (*ibid.*, l. 65)

(76) a sou.ffrir, et // qui plus me grie.(ve) (*ibid.*, l. 100)

From our point of view, these differ fundamentally from the Group I type of cæsura, and thus it is unfortunate that Guthrie combines statistics for both Group I and Group II cæsural types for the majority of the paper. Overall,

59. My thanks go to an anonymous reviewer of Rainsford (forthcoming) for highlighting the association of metrical and prosodic constituency.

60. Mid-14th century, in Hoepffner (1908–21).

61. In Guthrie's terms: 'the word in position 4 belongs to the major phrase in the second half of the line' (1987: 447).

Guthrie demonstrates that the proportion of lines containing a cæsura falls steadily from 81% in *Gormont et Isembart* to 60% by the 14th-century *Jugement dou Roy de Navarre*. At the same time, texts show a decline in iambic rhythmic organization: while in *Gormont et Isembart*, only 8% of stressed syllables occurred in weak metrical positions in the line, by the *Jugement dou Roy de Navarre*, this has risen to 27%. One of the causal factors, Guthrie suggests, is the 'progressive levelling of word stress' (1987: 67) at this time, a hypothesis which our discussion so far would support. Interestingly, Guthrie also demonstrates a difference in the versification of lyric and narrative verse throughout the time period.

> In lyric, and in the more experimental narrative line, the meter gradually becomes more abstract: the rate of fourth position cæsura declines [...] But the lyric is generally fifty to one hundred years ahead of the narrative line.
>
> (Guthrie, 1987: 64)

With the exception of innovative versifiers such as Chrétien de Troyes and Jean Bodel, narrative texts show a generally higher rate of fourth-position cæsura than lyric texts from a similar period. Guthrie argues too that even within the works of a single poet (Guillaume de Machaut), the lyric line shows more complex rhythms than the narrative line (1987: 68–69).

However, the extension of the notion of the cæsura to octosyllabic verse remains problematic. In ten- and twelve-syllable verse, the cæsura is clearly a metrical constraint, restricting a PhPh boundary to a particular position in the line.[62] Yet even with the inclusion of Group II lines, only 60–70% of lines in many texts show fourth-position cæsura. The possibility must surely remain that this is simply due to chance, and it is certainly not the case that metrical practice in the octosyllable is as regular as that of ten- and twelve-syllable verse.

4.4 Summary

The picture emerging from these studies is tantalizing but inconclusive. It is clear that the rhythm of verse changes from the earliest texts right up

62. Admittedly, some variations in cæsura position are observed: ten-syllable lines are generally divided into 4+6 syllables, but some 6+4 divisions are attested. However, there is a clear preference for 4+6. For example, in his detailed study of Jean Froissart's *Orloge amoureus* (1386; Dembowski, 1986), Billy (1999: 534) finds that only 4% of lines have a 6+4 structure, and this text is considered to be one of the more innovative of the ten-syllable verse tradition (1999: 543).

until the 14th century. It also seems that there is no discernable change in absolute metrical constraints; what changes can perhaps best be described as a 'rhythmic tendency'. Moreover, if the change is best described as the loss of iambic rhythm, as Noyer's (2002) work suggests, this is potentially very interesting for the present study. The rhythmic organization of word stress suggests that the opposition of stressed and unstressed syllables remains salient to speakers, which in turn suggests that stress deafness effects had not emerged. However, we must also consider the possibility that rhythmic organization is linked to a tendency to divide the line into two prosodic constituents. While this does represent rhythmic organization of a kind, it is less closely associated with the opposition of stressed and unstressed syllables at the word level, and thus cannot be so unambiguously linked to the emergence of stress deafness effects.

Chapter summary

In this chapter, I have discussed the features of the ModFr group-stress system, in particular rules establishing the position of primary and secondary stress. In a group-stress system, stress is regular at the level of the PhPh, a higher-level constituent on the prosodic hierarchy than the word. Moreover, I have argued that 'stress deafness' in the sense of Dupoux et al. (2001) is a inevitable correlate of such a stress system. Subsequently, I have examined phonological and morpho-syntactic changes attested in MedFr which have been linked to the emergence of group stress. Segmental phonological developments show that there is no unambiguous evidence for word stress beyond the 11th century. Morpho-syntactic developments such as the loss of V2 and the form of non-subject pronouns are not clearly stress-related. Finally, I have considered some evidence that the rhythm of verse texts develops throughout the MedFr period, and specifically that early texts may have shown a form of syllabo-tonic versification that I have argued to be incompatible with the stress deafness effects associated with group stress.

The conclusions drawn from this chapter shape the remainder of this thesis. In chapter three, I will conduct a thorough examination of rhythmic and syntactic constraints in verse texts, in order to develop a more precise chronology for the emergence of stress deafness effects. In chapter four, I will focus on the PhPh in MedFr, both to assess how it became the stress-bearing unit and what effect such a reanalysis may have had as it spread to other structures.

Chapter 2

Building a corpus

From the preliminary work on the character of group stress and the chronology of its emergence in chapter one, the rhythmic and syntactic development of verse (chapter three) and the development of the phonological phrase (chapter four) were identified as the core areas of research for the present thesis. The purpose of this chapter is to document the design of a corpus of texts that can be used for these studies. While the central focus of these studies will be to investigate change over time, it is equally important to control for factors other than diachrony which may influence the data, notably the verse form of the text, the dialect of the *langue d'oïl* area in which the text was written, and the type of text. In chapter three, the effect of these variables on the development of the rhythm of verse will be examined in separate studies, and therefore the corpus must contain a balance of texts in different forms, dialects and text types so as to permit meaningful analysis of this variation.

In recent years, a number of electronic corpora of MedFr texts have become available. For the extent of its coverage, Champion's proprietary *Corpus de la littérature médiévale* (Blum, 2001) is unsurpassed, containing 900 full texts in electronic format. [1] Only plain text searches are possible. Of the corpora which are freely available, the *Base de français médiéval* (BFM) (BFM, 2005) contains 26 full texts, searchable both as plain text and with part-of-speech tagging. [2] The *Nouveau Corpus d'Amsterdam* (NCA) (Stein et al., 2006) contains extracts of varying lengths from 200 texts mainly from the 13$^{\text{th}}$ century. [3]

1. Figure and full contents listed on the publisher's website: <`http://www.classiques-garnier.com/numerique-en/index.php?option=com_content\&id=47\%3Acorpus-of-medieval-literature`> [accessed 21 January 2011].

2. However, the majority of the texts are tagged automatically, using the freely available TreeTagger (Schmid, 1994), whose shortcomings are discussed in section 2.3.

3. Source: <`http://www.uni-stuttgart.de/lingrom/stein/corpus/`> [accessed 21 January 2011].

All extracts in the corpus have been manually annotated for part-of-speech and some morphological detail (see also §2.3). Most recently, the project *Modéliser le changement: Les Voies du français* (MCVF) has produced the first electronic corpus containing MedFr texts with structural annotation (see discussion in §2.5) (Martineau et al., 2010). Approximately 30 texts are included from the period up until 1500, of which five are administrative documents. [4] Work on the present study was well underway when this corpus was released.

The main advantage in basing a study on an existing corpus is the quantity of data that it is possible to search. For example, the BFM contains nearly one and a half million words, [5] while the NCA contains nearly three million words. [6] However, for the present study, there are substantial disadvantages. Firstly, the study would be limited to the texts included in the chosen corpus. Verse texts must be prioritized in the corpus design, and none of the existing corpora place a strong emphasis on having texts of a comparable form. Secondly, and more importantly, while these corpora are large, the annotation is too limited for the present study. If a number of studies are to be carried out on a large corpus of texts, machine-processable annotation is essential, both for corpus searches and to carry out statistical analyses. Studies of the rhythm of verse will require phonological and metrical annotation, and no current corpus is annotated in this way. Equally, it is clear from chapter one that group stress depends on the interaction of phonological and syntactic factors, and studies of the phonological phrase in chapter four will require both phonological and syntactic annotation. While the MCVF corpus is syntactically annotated, it contains comparatively few medieval texts, and it is not phonologically annotated.

Given the insufficiencies of existing corpora for the present study, it was necessary to construct a new corpus. Inevitably, this reduces the amount of data that can be searched simply on grounds of feasibility: existing corpora are all the results of major research initiatives. To give a preliminary size comparison, the final corpus used for this study contains 87 short extracts, each of approximately 3,000 words, of which only 27 are syntactically annotated. However, more elaborate analyses of the data are possible, and there is complete freedom over the choice of texts, thus a number of parameters of variation can

4. My estimate, based on the data provided at <http://gtrc.voies.uottawa.ca/fmi/iwp/res/iwp_home.html> [accessed 21 January 2011].

5. Source: <http://bfm.ens-lyon.fr/> [accessed 21 January 2011].

6. Source: <http://www.uni-stuttgart.de/lingrom/stein/corpus/> [accessed 21 January 2011].

be studied.

The chapter is structured as follows. In section 1, I discuss the selection of texts used in the corpus. In addition to the focus on verse, the corpus is structured to enable a comparison of different dialects and text types. Moreover, I discuss in detail the problems inherent in carrying out a systematic chronological study on source texts which cannot always be dated or localized accurately, have usually been recopied at a different time and place, and may even today be altered in the transfer from manuscript to printed edition. While the text selection aims to minimize these problems, a number of issues remain which must be borne in mind when the corpus is used in later analyses. In section 2, I provide a broad outline of the annotation scheme adopted in the corpus. Four 'layers' of annotation are discussed: syllabic, metrical, part-of-speech and structural annotation. The design of the syllabic and metrical annotation schemes is unique to this corpus, and requires a detailed understanding of medieval versification practice. Part-of-speech and structural layers are found in other corpora, but the merits of different systems are rarely discussed. In presenting the design of the annotation scheme for this corpus, I emphasize the similarities and differences between the present analysis and that of other corpora.

1 Selection of texts

In this section, I discuss the principles behind the selection of the corpus texts. In section 1, I present the core principles of the corpus design, which is divided into ten subcorpora, each containing texts from different periods which are comparable from the point of view of form, dialect and text type. Section 1.2 provides an overview of the texts used in the corpus. Section 1.3 focuses on the problems posed by the philological record and the steps taken to minimize their effect in the corpus design, while in section 1.4 I give an overview of the texts included in the base and prose subcorpora, on which a number of studies focus exclusively.

1.1 Division into subcorpora

The key feature of the corpus design is that it is split into ten subcorpora, and these are shown in table 1. Four variables are controlled for in the corpus design: date of composition, form, dialect and text type (narrative vs. theatre vs. lyric). The strongest emphasis in the corpus design is to select texts which

Subcorpus	Period	Form	Dialect	Text type
Early-8	1000–1175	8-syll	any	narrative
Base	1175–1500	8-syll	central	narrative
Anglo-Norman	1175–1300	8-syll	Anglo-Norman	narrative
Northern	1175–1300	8-syll	Northern	narrative
Eastern	1175–1300	8-syll	Eastern	narrative
Theatre	1150–1500	8-syll	any (pref. central)	theatre
Lyric	1150–1500	fixed forms	any (pref. central)	lyric
10-syll	1050–1500	10-syll	any (pref. central)	narrative
12-syll	1150–1500	12-syll	any (pref. central)	narrative
Prose	1150–1500	prose	any (pref. central)	narrative

Table 1: Subcorpora

are comparable in other ways, but composed at different times. There is one exceptional case: since the textual record is not sufficiently rich to include only texts in a central dialect from before 1175, the octosyllabic narrative subcorpus is split into two parts, 'early-8' (pre-1175, no dialectal controls) and 'base' (post-1175, central texts).

The division into subcorpora permits the studies in chapters three and four to compare, contrast or exclude particular kinds of variation while maintaining a balanced sample of texts in the corpus. Study 1 in chapter three compares the development of the rhythm of verse in different forms using texts from the early-8, 10-syll and 12-syll subcorpora. Study 2 in chapter three investigates the effect of dialect, contrasting the base, northern, eastern and Anglo-Norman subcorpus, and the effect of text type, using the base, theatre and lyric subcorpora. Study 3 focuses on the position of phonological phrase boundaries in lines of verse using only the base subcorpus, and excluding other types of variation. Studies of the development of the phonological phrase in chapter four use quantitative data from the octosyllabic narrative texts in the early-8 and base subcorpora, and from the prose texts. The studies in chapter four thus maintain a balance of verse and prose data in statistical analysis, with the other verse texts used qualitatively.

1.1.1 Form

The first variable used to distinguish the subcorpora is that of form. Studies of the rhythm of verse are only possible where the texts to be compared

have lines of the same length. The majority of subcorpora contain octosyllabic texts. The choice of the octosyllable as the 'default' form in the corpus reflects the frequency of its use in the medieval period: Zink, for example, argues that the octosyllabic rhyming couplet was 'the least marked literary medium' (1995: 50) before the emergence of prose romances in the second quarter of the 13th century, and it remains the most common verse form until the mid-16th century (Elwert, 1965: §162). Of all verse forms, the octosyllable is the only one to be used consistently in narrative, theatre and lyric versification throughout the medieval period. Moreover, from a rhythmic point of view, the octosyllabic line is the most interesting form to study, as the lack of a regular cæsura leads to rhythmic diversity. Contrasted with the octosyllable are two other common MedFr verse forms: ten- and twelve-syllable verse. Since lyric pieces show a variety of line lengths (sometimes within the same text), the form of the extracts is mixed, with seven-, eight-, and ten-syllable lines predominant. Finally, there is a prose subcorpus. The syntax of verse being artificially constrained by the metrical form, a prose subcorpus is essential for the more syntactic studies in chapter four, and also provides a control sample for the rhythmic studies in chapter three (study 4).

1.1.2 Dialect

The second variable used to distinguish texts is that of dialect. In the medieval period, the *langue d'oïl* area was not linguistically homogenous, comprising a number of dialects separately descended from vulgar Latin. ModFr is descended from a central variety of the *langue d'oïl*, primarily the dialect of Paris (Francian). [7] Literary texts of the the 12th and 13th centuries show dialectal features characteristic of the areas in which they were composed and copied. The complex manuscript traditions of many medieval texts mean that the text may be preserved in a manuscript copied in a different dialect area from that in which the text was composed. [8] In general, I have tried to avoid where possible any text where the author has been claimed to come from a different region from that of the scribe. Four subcorpora are used to study the effects

7. The term 'Francian' is a useful anachronism. The variety was known in the medieval period as *françois*.

8. The apparent 'mixing' of forms from different dialects within literary texts has even led some scholars to suggest that a supra-regional *scripta* existed in the 12th and 13th centuries (e.g. Gossen, 1967). As Lodge (1993: 114–15) points out, the clear regional differences in the written forms of both 13th-century legal and literary texts found by Dees (1980, 1987) make this theory problematic. In the present study, I assume that until the 14th century, texts which were both composed and copied within a particular region will reflect the dialectal features of that region.

of dialectal factors: the base subcorpus (central texts only), the northern subcorpus (Picard), the eastern subcorpus (texts from Lorraine, Franche-Comté, Burgundy), and the Anglo-Norman subcorpus (England). In these subcorpora, only those texts in which both the author and the scribe originate from the same dialectal region have been used.

The three non-central dialect areas were chosen because all were in contact with Germanic languages. In the case of the northern and eastern regions, it was here that Frankish settlement is argued to have been the most intense (e.g. Pope, 1952: §17), and the areas remained in close geographical proximity to Germanic-speaking regions. Since the strong tonic stress of the Gallo-Roman period is often attributed to Germanic influence (Pope, 1952: §223), the emergence of group stress may also have been affected by this continued language contact. Indeed, Pope (1952: §§1320–22) suggests a number of phonological developments which indicate that the lengthening of stressed vowels and the deletion or reduction of unstressed vowels were more extensive in the northern and eastern regions:

- Retention of falling diphthongs, e.g. /ˈie/ for central /jˈe/. Monophthongization by deletion of second element, e.g. *destrir* for central *destrier* (northern region) (Pope, 1952: §1320, iv).
- Early loss of schwa in hiatus with following vowel, e.g. *benois* for central *beneois* (northern region) (Pope, 1952: §1320, x).
- Reduction of unstressed /a/ to /ə/ in feminine determiners and pronouns, e.g. *me*, *le* for central *ma*, *la* (northern region) (Pope, 1952: §1320, xii).
- Diphthongization of stressed /ɛ/ and /ɔ/ in closed syllables, e.g. *apries*, *pierte* for central *apres*, *perte* (north-east and Lorraine) (Pope, 1952: §1321, iii).
- Diphthongization of tonic free /æ/, e.g. *meire* for central *mere* (north-east and eastern region) (Pope, 1952: §1321, iv).
- More rapid loss of final post-tonic schwa, unpronounced by the 15$^{\text{th}}$ century (north-east and eastern region) (Pope, 1952: §273; §1321, vi).

Anglo-Norman, on the other hand, is descended from a western dialect. Historically, Frankish influence in the west of France is claimed to have been less strong, and 'the most characteristic trait of its phonology in Old French is the relatively early levelling of diphthongs, a development that is probably indicative of a relatively early diminution of the tonic stress' (Pope, 1952: §1326). However, once introduced to England in 1066, Pope suggests that the stress pattern of English had a profound effect on the development of the di-

alect (1952: §1110). She suggests that the following dialectal developments may have been due to the influence of English stress:

- Loss of schwa in hiatus with a following vowel in the late 12th century, e.g. *emperur* for central *empereür* (Pope, 1952: §1132).
- Loss of final post-tonic schwa in all contexts from the 13th century (Pope, 1952: §1135).
- Deletion of initial syllables, e.g. *peler* for central *apeler* (Pope, 1952: §1137).

From our point of view, it is possible that prosodic change may have been slowed or even reversed as Anglo-Norman began to develop distinctly from continental French over the course of the 12th and 13th centuries.

The textual record does not permit a systematic study of the effect of dialectal variation before 1175.[9] However, it is also at this point that the dialect of the Île-de-France began to be perceived as the most prestigious variety of the *langue d'oïl*, and thus influenced written literary production in all dialects (Lodge, 1993: 97–102). Guernes de Pont Sainte-Maxence, for example, writing in *c.*1175, implies that the language of the Île-de-France is a 'good' variety ('Mis lenguages est buens car en France fui nez', *La Vie de Saint Thomas Beckett*, l. 6165, quoted in Lodge, 1993: 98–99). By the mid-14th century, dialectal differences in orthography are rare:

> Les différences dialectales disparaissent des textes vers le milieu du XIVe siècle, et le français règne dans tous les pays de langue d'oïl, avec sa nouvelle graphie.
>
> (Beaulieux, 1927: I, 152)

This is primarily true of literary texts, and more local documents such as charters retain dialectal forms in orthography for longer (Gossen, 1962). However, the emergence of a pan-regional *scripta* in the mid-14th century makes it more difficult to localize texts, and consequently the non-central dialectal subcorpora are limited to the late 12th and 13th centuries (1175–1300).[10]

1.1.3 Text type

The subcorpora distinguish three text types: narrative, lyric and theatre. Narrative is the default text type of the corpus, while lyric and theatre are

9. While there is no shortage of early 12th-century Anglo-Norman texts, there are few available central texts and virtually no eastern texts from this time.

10. This is similar to the period studied in Dees' (1987) comprehensive survey of dialectal forms in literary texts.

each represented by a single octosyllabic subcorpus.

The distinction between lyric and narrative texts is fundamental in the analysis of Zumthor (1972: 187ff.). Since a variety of narrative and lyric genres exist in the medieval period, no clear and consistent criterion distinguishing the two can be offered. However, a discourse-based distinction, owing much to Benveniste's (1966) categories of *récit* and *discours*, is particularly pertinent:

> Le discours personnel intègre, au moins fictivement, au texte, la présence de son énonciateur, et se réfère à un *hic* et *nunc*; il comporte, ne fût-ce que par là, une communication d'ordre fortement affectif. Le discours impersonnel se réfère au passé et comporte, quel que soit son contenu, une activité caractéristique : raconter.
>
> (Zumthor, 1972: 172)

'Impersonal' narrative discourse may also contain the first person singular *je*, but the narrative *je* represents the storyteller or author in the present, interacting with the audience or reader (Marnette, 1998). Zumthor (1972: 172–73) points out that first-person narratives are rare in MedFr, except in allegorical dream narratives.[11] Zumthor cautions against a simple association of personal discourse with lyric and impersonal discourse with narrative (1972: 175), but notes that such a distinction has considerable validity at the origins of French literature. Moreover, 'la survie de beaucoup de formes archaïques permet de conserver, pour l'essentiel [...], le même classement jusqu'au XIV^{e} et même XV^{e} siècle' (1972: 175). This consistency across the time period makes it feasible to incorporate a distinction of this kind into the corpus design.

Were the opposition purely discourse-based, it is less clear that it would be of relevance for a study of rhythmic change. However, there is also one vital rhythmic difference between narrative and lyric texts: in the lyric tradition, the text is subordinate to its musical setting, while in narrative, the setting is subordinate to the text. This is reflected in the form of the verse. Lyric texts are written in *strophes*: 'des entités [...] dans lesquelles une pluralité de vers se trouve réunie en un ensemble structuré' (Elwert, 1965: §177). The earliest lyric form to be attested in MedFr is the *chanson courtoise* of late 12$^{\text{th}}$-century *trouvères* such as Gace Brûlé and Conon de Béthune. Elwert emphasizes the effect of the musical setting on the verse:

> La strophe de la chanson courtoise de l'ancien [français] montre une articulation rigoureuse qui est conditionnée par le fait qu'elle se chantait

11. Joinville's *Vie de saint Louis* is a rare exception; other chroniclers (such as Clari or Villehardouin) refer to themselves in the third person (Zumthor, 1972: 173).

> sur un schéma musical défini.
>
> (Elwert, 1965: §194)

The earliest liturgical narrative texts from the 11th century too are sung, and written in regular *strophes*. However, over the course of the 12th and 13th centuries, narrative texts show progressively less complex settings. The earliest surviving *chansons de geste*, epic stories of the exploits of Charlemagne and his armies, date from the late 11th century. These too were sung, emerging from an oral tradition (Zink, 1995: 18). However, rather than the regular *strophes* of the lyric tradition, *chansons de geste* were composed in *laisses* of irregular length, united by a common assonance. They differ also from other short sung narrative texts (*chansons de toile*), which show both regular *strophes* and a refrain. For Zumthor (1972), this reflects a fundamental difference in the importance of the musical setting:

> On peut admettre l'existence de deux modèles extrêmes, en forte opposition l'un avec l'autre, et vers l'un desquels tend nécessairement chaque texte du corpus en question. D'une part, une variété de récit totalement assumé par le chant, comme le sont les paroles de la chanson de trouvère ; d'autre part, un récitatif servant simplement à soutenir le texte et à en varier le débit. [...] C'est donc de manière relativement paradoxale, mais non arbitraire, que je proposerai de distinguer, en gros, le *chant narratif*, d'une part ; le *récit chanté*, de l'autre.
>
> (Zumthor, 1972: 286)

Note here that 'la chanson de trouvère' is considered the extreme case in which the text is subordinate to the musical setting. *Chansons de geste*, however, are *récits chantés*: 'qu'une mélodie porte sans plus en constituer le lien ni le définir comme totalité' (1972: 311).

Moreover, from the mid-12th century, a new narrative genre and a new form emerged. The first romances were translations or retellings in the vernacular of classical histories (e.g. the *Roman de Thebes* tells the story of Œdipus), but they rapidly became associated with more original compositions (Bruckner, 2000: 13), most notably the elaboration of Arthurian legends. Formally, romances of the 12th century were written in continuous verse, using octosyllabic rhyming couplets. Unlike the *chansons de geste*, the romance had no musical setting. For Zumthor, the emergence of continuous texts is one of the most significant developments in medieval literature (1972: 339–40), and it is attested only in narrative. Narrative moves towards a form where it is 'libéré des contraintes du chant' (1972: 340), and formal constraints disappear entirely

in the 13th century with the development of prose romances. Yet the loss of musical settings in the mid-14th century did not result in a similar 'liberation' of lyric texts. Instead, a series of fixed forms emerged, notably the *ballade*, *rondeau* and *virelai* (Elwert, 1965: §§211–15; Zumthor, 1972: 270–71). The demands of the musical setting were replaced by the requirements of a rigid verse setting; moreover, rather than simplifying the verse form, 'les jeux de rimes et les jongleries de versification se compliquent et se multiplient' (Zumthor, 1972: 271). The distinction between narrative and lyric thus seems fundamental to any study of MedFr texts.

Theatrical texts develop later than either narrative or lyric, and initially as part of the liturgy; indeed, the earliest surviving play (the *Jeu d'Adam* from the mid-12th century) is interspersed with liturgical passages in Latin (Zumthor, 1972: 441–42). Formally, theatrical texts usually adopt the octosyllabic rhyming couplet of the romance, although a variety of line lengths may be used within a single text. In constrast to the formal changes attested in narrative and lyric texts, this form is used throughout the medieval period. The defining feature of the theatrical text is that it is written *par personnages*, with separate parts for individual characters, and is intended for performance. While lyric and some narrative texts too were performed, in theatre the performance element is even more fundamental:

> Le sens primaire de tout ensemble dramatique surgit de l'organisation d' 'acteurs' en un jeu stratégique et de 'personnages' dans une 'situation'.
>
> (Zumthor, 1972: 434)

Narrative, lyric and theatrical texts are each attested from the 12th century to the 15th century, although theatrical texts are rare in the 12th and 13th centuries. They are distinct both formally and in their context of communication. However, the classification poses two principal difficulties. Firstly, a number of texts and traditions are difficult to classify, a problem I will discuss in more detail in §1.4. Secondly, the 'narrative' text type encompasses a great variety of genres. Narrative includes sung texts intended purely for performance (e.g. the *chansons de geste*) and texts in prose destined for private reading (e.g. later romances).[12] While the division of subcorpora by verse or prose form will in part keep these texts distinct, it will be crucial to bear in mind the genre and context of performance of individual texts when the corpus data are analysed.

12. Although, even in private reading, the texts were spoken aloud (Zumthor, 1972: 38; Marnette, 1998: 212).

Date	Subcorpus									
	Early-8	Base	Theatre	Lyric	10-syll	12-syll	Prose	AN	Northern	Eastern
1000–	*PassClerm*^{ADP}				*Alexis*^{ADPM}					
	VieLeger^{ADPM}									
1100–	*Brendan*^{M}				*ChRoland*^{AP}					
	GormIsem^{ADPM}									
1150–	*Thebes*^{APM}				*CharNimes*^{APM}	*Rou*^{M}	*QuatrLivre*^{ADT}			
	LaisMarie^{M}									
1175–		*Charrete*	*JeuAdam*^{ADM}	*Blondel*^{M}		*Alexandre*		*Prothes*	*Eracle*^{M}	*Florimont*^{M}
				CononBeth^{ME}		*ChAntioche*^{A}			*BaratHaim*^{M}	
1200–			*JeuNicolas*	*GaceBrule*^{ME}	*AmiAmile*^{AP}	*AlexisO*^{AP}	*ConqVilleh*^{PM}	*VieEdmund*^{A}	*ChvBarisel*^{A}	*Dolopathos*
1225–		*MirNDCoin*^{M}		*ThibautCh*	*HuonBord*^{A}		*TristanPr*^{AP}	*GuiWarewic*^{A}	*ComtePoit*^{ADM}	*ImageMonde*^{A}
		PassJongl			*RaoulCamI*^{A}					
1250–			*MirTheoph*	*Rutebeuf*^{M}				*Chivalier*^{AD}	*BouchAbev*^{AM}	*IsopetLyon*^{AD}
				AdamHale						*Sacrist3*^{ADM}
1275–		*RoseMeun*	*Feuillee*^{E}			*Berte*^{M}	*MirLouis*^{M}	*VieRichard*	*RenartNouv*	*AbrCheval*
1300–		*ComteAnjou*	*PassPalat*^{AM}				*VieLouis*			
1325–		*LibFortune*^{AP}	*MirNDPers*^{A}		*Behaingne*					
1350–		*VoirDit*		*Machaut*	*OrlogeAmor*	*HugCapet*^{AM}	*Berinus*^{AM}			
						AlexisA^{AE}				
1375–			*Griseldis*^{AM}	*Froissart*			*ChrFroiss*			
1400–		*MutFortune*		*Christine*	*3Jugemens*		*AdvisionChr*			
		BelleDame		*Chartier*^{M}	*BrevNobles*		*QuadrInvec*			
1425–			*TroysGalans*^{AM}	*CharlesOrl*						
1450–		*Testament*	*PassGreban*	*Villon*^{M}	*LyonCor*^{A}		*JehSaintre*			
			Pathelin^{A}							
1475–			*JudithHolo*^{A}	*Molinet*	*EurialLucr*		*MemCommyn*			

* Key to philological problems: A = author, D = date of composition, P = place of composition, M = manuscript, E = edition, T = translation

Table 2: Composition of the ten subcorpora

Author	Subcorpus				
	8-syll narrative	Theatre	Lyric	10-syll	Prose
Jean Bodel	*BaratHaim*^{M}	*JeuNicolas*			
Adam de la Halle		*Feuillee*^{E}	*AdamHale*		
Rutebeuf		*MirTheoph*	*Rutebeuf*^{M}		
Guillaume de Machaut	*VoirDit*		*Machaut*	*Behaingne*	
Jean Froissart			*Froissart*	*OrlogeAmor*	*ChrFroiss*
Christine de Pizan	*MutFortune*		*Christine*	*3Jugemens*	*AdvisionChr*
Alain Chartier	*BelleDame*		*Chartier*^{M}	*BrevNobles*	*QuadrInvec*

Table 3: Texts by the same author

1.2 Overview of full corpus

Table 2 lists all 87 extracts in the corpus by subcorpus and by date. Names of the corpus extracts are given in abbreviated form: full titles, authors, date and place of composition, manuscript details, genre, edition and details of the extract used may be found in the bibliography. Superscript abbreviations are added to each text in these tables to denote philological concerns. The abbreviation A (author) denotes an anonymous text. The abbreviation D (date of composition) denotes a text whose date of composition cannot confidently be established to a period of less than half a century. The abbreviation P (place of composition) denotes a text whose place of composition cannot be confidently determined. The abbreviation M (manuscript) is used to denote problems in the relationship of the manuscript to the original text: a chronological separation of more than one hundred years between composition and manuscript (e.g. *Brendan*, *ConqVilleh*), or geographical origin for the manuscript different to that of the text itself (e.g. *JeuAdam*), or a text preserved in several different base manuscripts (mainly lyric texts, e.g. *Chartier*). The abbreviation E (edition) denotes the use of an edition with an unsatisfactory editorial policy (particularly *GaceBrule* and *CononBeth*; see §1.3.2). Finally, the abbreviation T (translation), applying only to *QuatrLivre*, indicates that the text is a translation. These problems will be discussed in more detail in section 1.3. Where possible I have included several works by the same author, and these are listed in Table 3. The clearest indicators of the effect of form and text type on language use may be found by comparing these texts, as differences due to chronology and dialect can here be assumed to be negligable.

1.2.1 Extracts

Extracts are generally taken from the beginning of texts, unless there are strong philological reasons to select from elsewhere (change of scribe, base manuscript, etc.). Extracts containing formulaic lists (e.g. of placenames or people) are avoided. For narrative and theatre, continuous sections of text are used, except in texts containing lyric pieces or prose splitting the narrative verse (e.g. *VoirDit*), in which case only the narrative verse has been retained. Where sections of a text are available in the part-of-speech tagged NCA, the selected extract is drawn from the extract in this corpus.

In the lyric subcorpus, the extract is made up of short pieces selected to ensure formally comparable extracts and a representative selection of the poetic forms employed. The approximately 500 lines of each extract are divided into

four components: three comprising poems written using seven, eight and ten syllable lines respectively, and a fourth component consisting of poems written using a mixture of line lengths.

1.3 Philological problems: pre-1150 texts as a case study

The corpus design requires texts to be classified according to date and place of composition, text type and form. This may seem to imply that we can give an exact value for each of these variables for each text in the corpus. This is not possible; however, I have attempted to choose texts and editions that are the least problematic for the criteria identified. In this section, I will focus on the pre-1175 texts in the corpus, which are among the most philologically problematic.

1.3.1 Date and place of composition

PassClerm (a passion narrative) and *VieLeger* (a saint's life) are the earliest texts in the corpus and among the earliest records of French.[13] They present unique philological difficulties: indeed, were it not for the fact they they are the only record of French from the early 11th century, they would not be included in the corpus. The manuscript containing the texts can be dated to *c.*1000, copied possibly in Clermont, possibly in Limoges (Avalle, 1962: 17). As both cities are on the edge of the *langue d'oïl* area, the manuscript shows a number of southern features, leading Linskill (1937: 143), in his edition of *VieLeger*, to attribute some forms to a Provençal scribe (e.g. *quandius*, l. 49). Basing himself on the linguistic evidence cited by Suchier (1878), Linskill (1937) argues that the text was written in the north-east of the *langue d'oïl* area. De Poerck (1963) not only rejects this hypothesis, but rejects the idea that linguistic evidence alone can be used to date or localize texts in a period for which there are no contemporary texts with a clear attribution against which they may be compared (1963: 3). De Poerck suggests instead that *VieLeger* is Poitevin in origin, arguing that the capitalization of the name of *saint MAXENZ* in the poem (l. 30) indicates that Maxence was an important local saint. For *PassClerm*, no external evidence is available (De Poerck, 1963: 12) and indeed Avalle (1962) uses mainly linguistic evidence to situate the text in Poitou.

13. Only three very short texts are older: the *Strasbourg Oaths*, a legal text from 842, the *Sequence of Saint Eulalia*, a short saint's life from the late 9th century, and the *Sermon on Jonah*, a badly damaged manuscript containing notes for a sermon written partly in Latin shorthand and partly in French (cf. De Poerck, 1955).

Finally, it is impossible to identify a date of composition for the texts, other than that they must have been composed before the manuscript was written in *c.*1000.

While *PassClerm* and *VieLeger* pose particularly acute problems of dating and localization, the other texts in the early corpus are also problematic. The majority are preserved in Anglo-Norman manuscripts (*Alexis*, *ChRoland*, *GormIsem*, *Brendan* and *LaisMarie*). However, of these only *Brendan* and *LaisMarie* are uncontroversially of Anglo-Norman origin.[14] *Brendan* is dedicated to the wife of King Henry I of England,[15] not only confirming the place of composition of the text but also allowing the date of composition to be fixed within a 25-year period at the beginning of the 12th century (Short and Merrilees, 1979: 4–6).

The origins of the *chansons de geste*, *ChRoland* and *GormIsem*, are more difficult to ascertain. *GormIsem* is anonymous, while the 'author' of *ChRoland* names himself in the last line of the poem as 'Turoldus'; however, it is not clear either who Turoldus was, nor whether he was the author, scribe, or performer of the poem (Moignet, 1969: 15–16). Indeed, since *chansons de geste* are associated with an oral tradition, it is not clear whether 'author' and 'performer' can be meaningfully distinguished.[16] In the case of *ChRoland*, the extant manuscript (from the second quarter of the 12th century) shows scribal errors consistent with having been copied, suggesting that an earlier written form of the poem existed (Short, 2005: I, 39). The original 'composition' of the text is dated from oblique references to more recent events (e.g. a reference to the camels and drums of the Saracens (l. 852) recalls the battle of Zaragoza in 1086; Moignet, 1969: 10). Rhyme and assonance are the most common source of evidence as to the dialect of the original poem; these give some indication of a western origin (e.g. non-dipthongization of tonic free /o/; Short, 2005: I, 93).

GormIsem is still more difficult to date and localize. It is a short text (only 660 lines) and represents only a fragment of a longer poem. Unlike *ChRoland*, which is preserved in a manuscript from the second quarter of the 12th century,

14. However, the author named in *LaisMarie* as 'Marie' is conventionally assumed to be the same as the author of a collection of fables; in the latter text, she refers to herself as being 'de France', thus of continental origin.

15. Although, as Short and Merrilees (1979: 4–5) note, there is some debate as to whether it is his first wife Matilda or his second, Alice.

16. It is a much debated issue whether the *chansons de geste* were composed by a single poet (as argued by Bédier, 1908–13) or were simply copied down based on an oral performance (Menéndez-Pidál, 1959; Duggan, 1973); cf. Zink (1995: 25–32); Berthelot (2006: 32).

the only surviving manuscript dates from the 13th century, and thus cannot provide a useful *terminus ad quem* for the composition of the text. Little internal evidence for dating the text is available; in terms of external evidence, a performance of a version of the poem is recorded in the chronicle of Hariulf at Saint-Riquier at the end of the 11th century (Bayot, 1931: x). Whether the preserved fragment contains this version (as argued by Calin, 1961), or whether it dates instead from early in the 12th century (cf. Paris, 1902), it is difficult to fix the date of composition with any precision. Fixing the place of composition is so problematic as to be virtually impossible. Bayot (1931) argues that once the Anglo-Norman scribal forms have been 'removed', 'il reste un texte qui ne caractérise aucun trait dialectal notoire' (Bayot, 1931: v). More positively, Pope (1918) highlights a number phonetic and morphological features characteristic of dialects to the south-west of Paris. However, 'removing' the Anglo-Norman scribal forms from the manuscript before dialectal analysis instantly dismisses the possibility that the original text could also have been Anglo-Norman. From the point of view of a linguistic study, very little dialectal information can be gained from texts of this kind where the dialect of the author and that of the scribe is argued to differ.

Finally, *Alexis* (a saint's life) is impossible to date or localize with any accuracy. While the manuscript dates from the first half of the 12th century, the linguistic features of the text suggest that it was composed before *ChRoland* but after *VieLeger* and *PassClerm*. Paris and Pannier (1887) thus date the text to the mid-11th century. As it is not feasible to review debates on dating for every text in the corpus, I generally use the date adopted by the bibliography of the *Dictionnaire étymologique de l'ancien français* (DEAF) (Möhren, 2007), a practice also followed by the NCA and the BFM. In the case of *Alexis*, this is the late 11th century. With no texts of a comparable age, it is very dubious whether it can ever be established where in France the original text was written (Paris and Pannier, 1887: 43). [17]

Despite the problems of dating and localization, we have no choice but to use these early texts if we are to draw any conclusions about MedFr from before 1150. However, for periods in which there is a greater choice of texts, I have attempted to select those whose dating and localization, like that of *Brendan*, is rather clearer. In particular, it is useful to include texts where something is known of the author, either because several works are attributed to them (e.g.

17. Paris and Pannier (1887: 43–46) speculate that it was written in Normandy by the monk Tedbalt de Vernon; however, as Storey (1968: 22–23) points out, there is little evidence for this claim.

Chrétien de Troyes, Adam de la Halle), or because they can be identified with a historical figure (e.g. Geoffroy de Villehardouin, Jean de Joinville).

However, there remain a number of texts in all subcorpora with problematic dating and localization. While I have generally avoided texts that cannot be dated to within 25 years, an exception was made for two 13th-century *fabliaux*:[18] *Sacrist3* and *Chivalier*. *Chivalier* is one of the very few Anglo-Norman *fabliaux*, while *Sacrist3* was included for its Eastern attribution.[19] *IsopetLyon* too cannot be accurately dated, but can be clearly localized in Lyons, and hence was included in the eastern subcorpus. *JeuAdam*, the earliest play, and *QuatrLivre*, one of the earliest examples of non-legal prose, are also included despite only being datable to within half a century.

1.3.2 Transmission and editorial policy

While of the early texts *Brendan*, *LaisMarie* and *Thebes* are the least problematic with regard to dating and localization,[20] all are preserved in manuscripts copied in the 13th century, over a hundred years after the texts were written. A substantial gap between the date of composition and date of manuscript is potentially problematic if revisions to the text are carried out by the scribe: the result being a chronological hybrid text of limited use to the present study.

Of the early texts, *Thebes* would seem to have been the most affected in this way, a point highlighted in the DEAF bibliography. Raynaud de Lage (1966) dates the text of his 13th-century base manuscript to *c.*1160, based on comparative literary evidence. However, he observes that the text has been 'partiellement récrit' (1966: xv), with some modernization of the language. An older state of the text is preserved in the much later 14th-century Anglo-Norman manuscript (British Library Add. 34114, Ms. S), which is edited by Mora-Lebrun (1995). This manuscript too, however, is not without modifications: a third of lines show a syllable count which does not conform to the

18. Short, often bawdy narratives.

19. It is marginal whether this text should be included. In favour, based on linguistic evidence and geographical references in the text, the editors are certain of a Burgundian origin (Noomen and van den Boogaard, 1983–98: VII, 8), and it proved a difficult task to find samples of eastern octosyllables from the 13th century. However, there is almost no evidence permitting the text to be dated: Noomen and van den Boogaard (1983–98: VII, 7) rely on the loss of the case system to suggest a date within the 13th century. The manuscript dates from the end of the 13th century.

20. *Thebes* originates in the south-west, as shown by linguistic traits in both the base manuscript of Raynaud de Lage's (1966) edition and in an earlier fragment preserved in a manuscript copied between 1180 and 1220 (Raynaud de Lage, 1966: vi; xvii).

continental norm of the octosyllable (1995: 34). Mora-Lebrun attributes this to the Anglo-Norman scribe's treatment of schwa rather than to the original poem, and indeed the manuscript published by Raynaud de Lage is much more regular. If this manuscript is truly a chronological hybrid, it is questionable whether assigning it a date of composition of *c.*1160 (as both Noyer (2002) and the NCA bibliography do) is justified. In the event, I have done so with reservations. The problems with this dating must, however, be considered when data from the text is analysed.

The text as presented in modern critical editions is rarely the unaltered text of a single manuscript. However, in some cases, the text of critical editions is far removed from any of the source manuscripts. Both *GormIsem* and *Alexis* are preserved in Anglo-Norman manuscripts but are argued not to be of Anglo-Norman origin. Paris and Pannier (1887) (*Alexis*) and Bayot (1931) (*GormIsem*) produce critical editions of these texts with the Anglo-Norman features removed, and the syllable count, orthography and case system regularized. Such texts are a modern fiction rather than a historical reality of any kind,[21] and cannot be used in a linguistic study.[22] More appropriate are editions which reproduce a single manuscript with a minimum of editorial intervention, and in which those interventions are clearly marked: thus, for example, I follow Marnette (1998) and the BFM in using Moignet's (1969) edition of the *Chanson de Roland* rather than the more interventionist editions of Segre (1989) or Short (2005). However, where editorial interventions are made, I have followed the reading of the modern editor rather than reinstating manuscript variants in the corpus. An advantage of using a text with some editorial interventions is that the most obvious scribal errors are removed. Clearly, these can and do occur (e.g. repeated words, eyeskip causing omitted lines, etc.) and where they occur, they disrupt the sense of the text, making syntactic annotation more difficult. However, in the analysis of individual tokens from the corpus where morphological form is important (cf. chapter four, §2), I refer back to the printed edition to check both the manuscript reading and attested variants.

In selecting an edition, I was guided both by the editorial policy given

21. Even Paris acknowledges that what he has created is not an edition of the manuscript, but 'un spécimen admissible de la bonne langue française telle qu'elle devait se parler et s'écrire au milieu du XI^{e} siècle' (Paris and Pannier, 1887: 135). For a discussion of the methods of reconstructive editors such as Gaston Paris and his followers, see Cerquiglini (1989: 73–104), who exemplifies his discussion with Paris's edition of the *Vie de saint Alexis*.

22. For *GormIsem*, the diplomatic transcription of Bayot (1931) is used, not the critical edition.

in the introduction and by the editions selected for the NCA and the BFM. Preference was given to editions reproducing a single base manuscript with clearly marked variants. As the studies in chapter three focus on the form of verse, I was careful to avoid editions in which I was aware that the editor had consistently intervened to regularize the syllable count of the manuscript. In *LibFortune*, where the only available edition (Grigsby, 1967) contains this type of intervention, the reading of the base manuscript as given in the variants was used rather than the editor's text.

Editions of lyric verse often fall short of the single manuscript ideal, varying the base manuscript piece by piece. In most cases, this is because the collection is a modern creation, and no base manuscript preserves the full corpus of the poet's work. The editions of *Blondel*, *ThibautCh*, *Rutebeuf*, *AdamHale*, *Chartier* and *Villon* all fall into this category. The selected editions of *CononBeth* and *GaceBrule* go still further, creating a composite text based on all source manuscripts, with a 'base' manuscript used only to provide orthography. In both cases, these editions have been used only out of necessity. For *CononBeth*, Wallensköld's (1921) edition is the only one available. For *GaceBrule*, both Dyggve's (1951) edition and the earlier edition of Huet (1902) are comparative. Rosenberg and Danon (1985) provide the only non-composite edition of the poet's work; unfortunately, I was not aware of its existence at the time of compiling the corpus.[23]

1.4 The base and prose subcorpora

Data from the base and prose subcorpora are used in the majority of studies, and a few preliminary remarks about the texts included are required.

QuatrLivre is a translation of the two books of Samuel and the Kings from the Latin Old Testament. The potential influence of Latin, in particular regarding word order, has led to translations elsewhere being excluded from the corpus.[24] However, Herman (1990: 260–61) argues that comparison with the Latin source shows that the translator does not slavishly follow the lexis or word order of the original; in fact 'l'ordre des termes essentiels de la proposition est pratiquement indépendant de l'ordre latin' (1990: 261). This suggests that

23. A subsequent comparison of this edition with Dyggve's critical text revealed the differences to be mainly morpho-phonological. Dyggve's version as used in the corpus is thus suitable for the analyses of verse rhythm in chapter three, but is used with caution in other studies.

24. *AbrCheval*, a *mise en vers* of a prose translation of Vegetius' *De re militari*, was considered far enough removed from the original to be included.

the text may be used as a source of MedFr data with a degree of confidence. Moreover, the rarity of 12th-century literary prose means that the choice of texts is limited.

As with the *chansons de geste*, *PassJongl* has an oral tradition, and would seem to have been performed by *jongleurs* since the end of the 12th century or the beginning of the 13th century (Perry, 1981: 18). Thus, while the text used is drawn from the *Bible des sept estaz du monde* compiled by Geufroi de Paris in 1243, [25] it is slightly debateable whether this is its date of 'composition'. Certainly, we may expect it to contain more conservative linguistic features than other mid-13th-century texts. *PassJongl* is also 'semi-dramatique' (Perry, 1981: 24). While the text is intended for a single performer, Perry argues that the performance may have been highly mimetic, and thus, in many ways, already similar to the theatrical passion plays of the early 14th century such as *PassPalat* (1981: 24–26). The inclusion of later passion plays in the theatre subcorpus, including the closely related *PassPalat*, makes the inclusion of this 'semi-dramatic' piece appealing. As it is not clearly composed as a play (with character's 'parts' marked outside the body of the text), I have elected to classify it as narrative. [26]

RoseMeun I have chosen to classify as a central text despite the fact that Jean de Meun originated from near Orleans (e.g. Lecoy, 1965–75: vi). The language of the text is not as markedly Orléannais as the first part of the *Roman de la Rose* written by Guillaume de Lorris (DEAF). Moreover, the language of the contemporary base manuscript used by Lecoy (1965–75) shows clear Francian traits: Dees (1987: 527) localizes the manuscript in the Paris region. [27]

VoirDit and *BelleDame* lie on the boundary between narrative and lyric. Zumthor (1972: 309–10) classes *VoirDit* and *BelleDame* as *romans-en-poésie*:

25. The date has been contested (see Perry, 1981: 95–96), but even if a more approximate date of 1243–63 is proposed, this is still precise from the point of view of this corpus.

26. A similar distinction is drawn by Jodogne (1965): the text lacks clear 'impersonation' (cited by Perry, 1981: 25).

27. Dees (1987) makes a valuable contribution to the problem of localizing manuscripts, and where available, I have made use of his data in designing the corpus. Having established the morpho-phonological characteristics of different regions of the *langue d'oïl* based on the orthography of legal texts whose localization was clear (Dees, 1980), Dees then uses these characteristics to localize (mainly 13th-century) literary manuscripts. For each text, a coefficient indicates the certainty of the localization (see Dees, 1987: 519–33). I have made a point of including texts with a particularly high dialectal coefficient (over 90), showing especially clear dialectal features. However, it should be noted that Dees's localizations apply only to the text as preserved in an individual manuscript, and the distinction between scribal dialect (shown by morpho-phonological forms) and authorial dialect (shown by rhyme or assonance) favoured within the philological tradition is not considered.

both have a *je* narrator, but incorporate elements inherited from the lyric tradition (e.g. the debate between lover and lady which constitutes the major part of *BelleDame*). Nevertheless, even in the 15th century there remains an opposition between these longer texts, which incorporate narrative elements, and shorter lyric pieces in a fixed form: witness the inclusion of lyrics in *VoirDit*. For this reason, I have chosen to classify these texts as narrative rather than lyric. *Testament* too proves difficult to classify within the typology adopted: while there is no strong narrative element, lyric pieces are interspersed within the octosyllabic *huitains* which form the body of the text. While the atypical form of the text might have constituted a reason for excluding it from the corpus, it is retained in the base subcorpus since, as perhaps the only medieval text for which there exists a complete metrical commentary (Pensom, 2004), it is interesting to see whether its metrical and rhythmic features are also unusual.

1.5 Summary

In this section, I have outlined a corpus design which permits similar texts from different points in the medieval period to be compared. The texts are distinguished by three parameters: form, dialect and text type. However, recognizing that any classification over-simplifies the complexities of the philological record, I have attempted firstly to select texts which are least problematic from a philological point of view and secondly to highlight problematic texts from the outset, so that these difficulties are considered in the analyses in chapters three and four.

2 Annotation

The second major challenge in building the corpus was to design and implement a machine-processable annotation scheme. This is required for a number of reasons. Firstly, aspects of the stress system must be inferred from lexical, syntactic and morphological properties, and machine-processable annotation ensures that these inferences are made consistently. Secondly, it is essential for large-scale quantitative studies, such as those in chapter three. Thirdly, it reduces the time required to a locate a given phenomenon within the corpus texts, increasing the flexibility of the corpus for the diverse studies in chapter four. The purpose of this section is not to provide an annotators' manual, nor a corpus user-guide. Instead, it presents an overview of the assumptions made by

the analysis encoded in the annotation, and relates the scheme adopted in the present corpus to that of existing corpora. An understanding of the analysis adopted here is not essential for the interpretation of the results of the studies in chapters three and four.

The corpus annotation is divided into four layers. Syllabic annotation consists of labelling the number and type of syllables in each orthographic form (§2.1). Metrical annotation maps the syllables in the text onto metrical positions in a line of verse, and also marks features of the line itself (e.g. length, regularity of scansion, cæsura type, etc.) (§2.2). The third layer is part-of-speech annotation (noun, verb, pronoun, etc.) (§2.3). The fourth layer is structural annotation, which encodes the hierarchical syntactic structure of the text (§2.5). All layers are machine processable. The annotation can thus be used either to search the corpus for individual tokens, or as the basis for quantitative studies. More sophisticated analyses are also possible, such as the reconstruction of stress position and of phonological phrase edges in lines of verse (cf. chapter three, §1).

Syllabic and metrical annotation is not found in any published corpus of French, medieval or modern, and the design is devised from first principles. Part-of-speech and structural annotation have been developed for many other corpora, including MedFr corpora, and the design adopted in this corpus draws on previous studies. Details of the annotation adopted in existing corpora are discussed in the introduction to section 2.3 and in section 2.4; the solution adopted to syntactic annotation in the present corpus is presented in section 2.5.

2.1 Syllabic annotation

2.1.1 Overview

Syllabic annotation marks the number and type of syllables in each word. [28] The syllabic annotation is essential both as an input to the reconstruction of prominent syllables (chapter three, §1.1) and as an input to the metrical annotation.

28. The notion of 'word' is problematic in the light of varying conventions of word division both in manuscripts and modern editions. I return to the issue in sections 2.3 and 2.5.

2.1.2 Concordance

Syllabic annotation was carried out entirely manually, with no machine assistance. A central concordance of all orthographic forms in the corpus was compiled, and each form was tagged. The tag in the concordance was then automatically applied to each occurrence of the word in the corpus. Thus only 30,000 orthographic forms had to be annotated, rather than all of the approximately 250,000 words in the corpus. The concordance approach has two main advantages: it reduces the time needed to annotate the full corpus, and consistency of tagging across the corpus is ensured.

However, there are also disadvantages. In the five hundred years covered by the corpus, some syllables are lost, in particular schwa in hiatus with a following vowel (see below). The annotation must provide a means to model this change. A more serious problem is the inconsistency of MedFr orthography. The most common type of inconsistency is variable orthography for a single phonetic form (e.g. <kar>, <quar>, <car>; now ModFr *car*). This type of variability poses no problem to a concordance: each form is included and given an identical tag. However, in cases where a single orthographic form may represent two or more different words with different syllabic forms, the concordance approach encounters problems.

The solution adopted is to store two orthographic forms of each word in the corpus. One represents exactly the form used by the modern editor, and it is this form that is visible in results derived from the corpus. The second form, used by the concordance, contains minor modifications to the editorial form (e.g. no capitalization, no parentheses or brackets to denote letters added or removed from the manuscript form) and standardizes the use of diacritics. Three cases in particular may be distinguished in which the standardization of diacritics is essential to the concordance approach. Firstly, there is the case described above where linguistic change causes a schwa in hiatus to be deleted in later texts: [29]

(1) <veoir> /vəˈwɛr/ (most 14th- and 15th-century texts)
/ˈvwɛr/ (*TroysGalans*)

Modern editors would usually use a diæresis (¨) to denote the disyllabic form (i.e. *vëoir*), and no diacritic in the monosyllabic form. In this case, where

29. Pope (1952: §§268–69) and Fouché (1952–69: 516) date this change as beginning in the 13th century and being completed by the 16th century. My own corpus confirms that the greatest variation in the scansion of these syllables is found in the early 15th century; cf. §2.2.

the lexemes are identical and only the syllable count of the line determines whether the form is monosyllabic or disyllabic, I prefer to label the syllable which is deleted in later texts specifically, and to use the metrical annotation to specify either the monosyllabic or disyllabic form. In the concordance, the editorial diacritic is eliminated, and I allow for both possible realizations in (1) in the annotation of the single form <veoir>. Secondly, there are cases in which different scribal conventions cause two distinct words to be written identically:

(2) <armez> /arˈmæs/ (most 14th-century texts)
/ˈarməs/ (*HugCapet*)

Here, the scribe of *HugCapet* uses <-ez> for final /-əs/ rather than just for final /-ˈæs/, as is most common in other texts. Thirdly, there are cases of homographs found in all texts:

(3) <oir> /ˈwɛr/ 'heir' or /oˈir/ 'to hear'

Most editors use a diæresis to distinguish the monosyllabic reading 'heir' from the disyllabic 'to hear' (i.e. *oir* vs. *oïr*). I adopt the same convention in the concordance, except that in keeping with the practice adopted for (1) I retain the diæresis even in cases where <oir> 'to hear' is monosyllabic. These minor modifications to diacritics ensure a many-to-one, rather than a many-to-many, relationship between orthographic and phonological form, essential for a concordance-based approach.

2.1.3 Tagset

The syllabic annotation must provide sufficient information to reconstruct the position of prominent syllables (cf. chapter three, §1.1). Recall the simple MedFr word stress rule from chapter one:

> **MedFr stress rule**
> Stress the final syllable of the word, unless the vowel in that syllable is a schwa, in which case stress the penult.

Two pieces of information are required to identify the position of primary stress within a word:

(a) Number of syllables
(b) Nature of the final syllable

Tagging is extremely simple: `v` is used to denote a non-schwa vowel and `s` to denote a schwa. This is exemplified with the following extract from the concordance, containing forms of *avoir*:

ai (1sg pr. ind.)	/ˈai/	`v`
aie (1sg pr. subj.)	/ˈai.ə/	`vs`
avons (1pl pr. ind.)	/a.ˈvõns/	`vv`
avoie (1sg imp. ind.)	/a.ˈvwɛ.ə/	`vvs`

From these tags, it can be correctly deduced that *ai* is a monosyllable, *avons* is an oxytonic disyllable, *aie* is a paroxytonic disyllable, etc.[30]

A complication arises in cases such as (1) above, where linguistic change causes a syllable to be deleted. Schwa in hiatus with a following vowel is explicitly tagged as `h`:

veoir	/v(ə.)ˈwɛr/	`hv`
conreé	/kõn.r(ə.)æ/	`vhv`
coneue	/kõ.n(ə.)y.ə/	`vhvs`

The `h` tag allows these syllables to be treated by a specific deletion rule in 15th-century texts when the metrical annotation is generated (see below, §2.2). The tag is only used for word-internal schwa in hiatus, not for word-final schwa which may be elided before a following word-initial vowel.

A second area in which linguistic change affects the syllabic structure of words is the treatment of diphthongs. In MedFr, there are two main historical sources for adjacent vowels: diphthongization of a single etymological vowel (cf. chapter one, §3.1) and two separate etymological vowels brought into hiatus by the disappearance of an intervocalic consonant (Pope, 1952: §§224–33, §237; Fouché, 1952–69: 219–27, 437):[31]

(4) AUDĪRE > */ɔ.ˈðir/ (10th century) > /o.ˈir/ <oir> (12th century)

(5) HĒREM > */ˈeir/ (10th century) > /ˈoir/ <oir> (early 12th century)

The loss of intervocalic /ð/ in (4) causes two etymologically separate vowels to be in hiatus, while in (5), the /oi/ diphthong is derived from the breaking of primary stressed Ē in an open syllable. The difference in pronunciation of the two vowel sequences (/o.ˈi/ vs. /ˈoi/) is shown both by the syllable count of medieval verse and the subsequent development of the forms. In this case, the vowels in hiatus never merge phonetically with the diphthong: developments of the former remain as ModFr glide plus vowel /wi/ (e.g. *inoui*), while the latter develops through several stages to ModFr /wa/.[32] Such cases are unproblem-

30. Phonetic transcriptions are given for reference and are not included in the concordance.

31. Diphthongs may also be formed by the combination of a single etymological vowel with a semi-vowel /j/ or /w/ derived from a number of sources; see Pope (1952: §506). From the point of view of the current discussion, these develop in the same way as diphthongs derived through diphthongization of a single vowel.

32. The development of vowels in hiatus is complex (Pope, 1952: §§237–47). Cases in which the vowels remain separate (1952: §§238–39) are unproblematic for the present study: they

atic for syllabic tagging: vowels in hiatus may be tagged `vv` while diphthongs may be tagged `v`.

However, in a number of cases, the pronunciation of vowels in hiatus and diphthongs does converge. Lote (1949–96: III, 127) notes some single syllable scansions of vowels in hiatus, although there is considerable lexical variation. For example, vowels in hiatus in *moelle*, *coette* and *boette* are regularly scanned as one syllable in the 15th century, as is *oïl*, first attested as a monosyllable in the work of Adam de la Halle (last third of the 13th century). Yet in *poete*, *noel*, *rouelle* and others, the vowels in hiatus are retained even in the 15th century. More troublesome still is the sequence written <ie>. Unlike the <oi> sequences, the quality of the diphthong and the vowels in hiatus remain identical throughout the medieval period. Moreover, vowels in hiatus and diphthongs occur in very similar morphological contexts:

(6) **Final <-ier>**
   (a) -A.RUM > /-ˈiær/ (e.g. *denier*, *destrier*, *tablier*)
   (b) -I.TĀ.RE > /-i.ˈær/ (e.g. *oblïer*, *marïer*)

(7) **Final <-ien>**
   (a) REM, BENE > <rien, bien> /riɛ̃n, biɛ̃n/
   (b) -I.Ā.NUS > /-i.ˈɛ̃n/ (e.g. *chrestïen*, *ancïen*)

(Lote, 1949–96: III, 123–24)

Both <-ier> endings are suffixes, and the two <-ien> endings are both word-final. Merger of the two vowel sequences in pronunciation may have been favoured by the similar contexts in which they occurred. Lote observes examples of both one and two syllable scansions of the vowels in hiatus throughout the medieval period (though more commonly in the 15th century), suggesting that 'les poètes ne se font aucun scrupule de compter tantôt d'une manière, tantôt de l'autre' (1949–96: III, 125). Moreover, occasionally the etymological diphthongs show a disyllabic scansion, particularly at the rime: Lote gives examples such as *bi.en*, *vi.erge*, *chi.en*, even *Di.eu* from 15th-century texts, concluding that 'les poètes se permettent de prendre les libertés les plus inattendues' (1949–96: III, 129). Alternatively, this could be seen as evidence that the systematic difference in pronunciation between vowels in hiatus and diphthongs of the 12th century had disappeared, and consequently poets could no

are tagged as separate vowels (`vv`). Cases in which the vowels merge into a monophthong (1952: §§242–45) are dealt with using the `h` tag. The cases dealt with in the following paragraph, however, are those in which the first vowel 'consonantalizes': i.e. in Pope's (1952: §241) view, becomes a semi-vowel /j/, /w/ or /ɥ/. In this case, two distinct vowel qualities are retained, but they receive only a monosyllabic scansion, exactly as do diphthongs derived from single vowels.

longer distinguish between those cases in which <ie> historically formed a hiatus and those in which it did not.[33]

From the point of view of syllabic annotation, a specific tag is used to denote those vowels in hiatus (in particular <ie>) which may adopt a monosyllabic pronunciation in later texts:[34]

ancien	`vdd`
celestien	`vvdd`
oublier	`vdd`

When metrical annotation is carried out, a rule determines whether `dd` groups should be counted as one or two syllables.

2.2 Metrical annotation (scansion)

Metrical annotation associates individual syllables in the text with metrical positions in lines of verse. The annotation represents the text as a series of lines, adding tags to each line to show its syllable count, cæsura type, ending type and any specific elision rules applied. In this section, I use the equivalent term 'scansion' in preference to 'metrical annotation'.

Annotating the scansion of texts manually is not feasible: a single extract contains around four thousand syllables. Instead, a customizable scansion algorithm is used, taking the syllabic annotation, the position of line breaks and a series of text-specific scansion rules as its input. This is not a simple task. While modern philologists have established 'rules' of MedFr scansion, these are not exceptionless, even in texts which are supposedly metrically regular. The algorithm applies a number of standard and some non-standard elision rules in order to produce a plausible regular scansion of a line of verse. The algorithm takes a flexible approach to scansion, in order to deal with the variability in the medieval texts.

2.2.1 Meter of French verse: standard rules

Before discussing the algorithm, I will begin with a more detailed look at the rules suggested for MedFr scansion by Lote (1949–96) and Elwert (1965) (cf. also chapter one, §4). The syllable count of lines is complicated by (i) the

33. Nor is the trend restricted to the 15th century: the poet of *PassPalat*, writing at the beginning of the 14th century, regularly adopts a hiatus scansion of the diphthong in *denier*.

34. I.e. <-ien> and <-ier> endings as above; <-ion> endings in Latin borrowings (from -A.TI.Ō.NEM); <-ieus> endings borrowed from -I.Ō.SUS (e.g. *glorieus*); and a number of specific lexical items (e.g. <ia> in *diable*).

status of schwa, (ii) the treatment of diphthongs, and (iii) the nature of the cæsura. The following is a synopsis of the key points:

(i) Line-final schwa syllables are pronounced but never counted ('feminine' as opposed to 'masculine' rime) (Elwert, 1965: §31).

(ii) Within a line, word-final schwa in a closed syllable (i.e. <-es>, <-ent>) is pronounced and counted (Elwert, 1965: §37).

(iii) Within a line, word-final schwa in an open syllable is pronounced and counted before a word beginning with a consonant (Elwert, 1965: §32).

(iv) Within a line, word-final schwa in an open syllable is either linked (written but not counted) or elided (not written) before a vowel (Elwert, 1965: §33).

(v) Word-internal schwa counts as syllabic (Elwert, 1965: §37), unless in orthographic proparoxytones (e.g. <angele>, <imagene>; see chapter one, §3.1.1).

(vi) Diphthongs count as a single syllable only if they stem from a single syllable etymologically (Elwert, 1965: §38).

To illustrate these rules in practice, consider a short extract from *Charrete*:

(8) A.prés la bie.re ve.nir voi.(ent)
u.ne ro.te‿et de.vant ve.noit
uns granz che.va.liers qui me.noit
u.ne be.le da.me‿a se.nes.(tre.)

(*Charrete*, ll. 54–57)

In (8), each line consists of eight metrical positions with no cæsura. The non-counting of final schwa syllables (rule i above) is illustrated in *voient* and *senestre*. The rule applies both to closed and open syllables. Elision[35] of final schwa in an open syllable before a following vowel (iv) is shown by the non-counting of the final syllable of *rote* and *dame*. Elswhere, open schwa syllables are counted, whether word-final (*biere*, *une*, *bele*) or word-medial (*venir*, *devant*, *chevaliers*, *menoit*, *senestre*). The five two-vowel sequences in this extract (<ie> in *biere*, *chevaliers*; <oi> in *voient*, *venoit*, *menoit*) are all diphthongs rather than vowels in hiatus,[36] and thus count for only one syllable by (vi)

35. Elwert (1965: §33) claims that 'linked' schwa are pronounced, but not counted, noting a similar phenomenon in which pronounced post-tonic full vowels are not counted before a vowel-initial word in Spanish and Italian verse. Whether or not this is the case, I assume that it makes no difference to the rhythmic structure of verse, and will henceforth refer to all non-counting of schwa before a vowel as 'elision'.

36. All are descended from a single etymological vowel, vulgar Latin /a/ in *chevaliers*, vulgar Latin /e/ in *venoit* and *menoit*; the source of *biere* is Frankish **bëra*.

above.

The syllabic tag detailed in §2.1.3 provides enough information to model all but elision of schwa across word boundaries (iv above). Two further pieces of information are required in the concordance of orthographic forms:

(c) Is the form vowel-initial? (y/n)

(d) Does the form contain a word-final open (elidable) schwa syllable? (y/n)

With these parameters included in the concordance, the scansion algorithm can elide word final schwa in contexts where rule (iv) applies.

Ten- and twelve-syllable lines contain a cæsura (cf. chapter one, §4.3). As at the end of the line, the final syllable before this cæsura is typically stressed (Elwert, 1965: §85); this is a masculine cæsura:

(9) or en.ten.**dez** // sei.gnor gen.til ba.ron (*AmiAmile*, l. 1)

If a word-final schwa syllable falls after the cæsura position, it may be discounted, as at the end of the line (epic cæsura):

(10) que deus de **gloi.(re)** // voz fa.ce vrai par.don (*AmiAmile*, l. 2)

The name derives from the fact that this type of cæsura is most common in epic texts, mainly *chansons de geste* (Elwert, 1965: §86). In lyric texts, word-final schwa may usually only fall after the cæsura if elided. Rarely, it may count as the first syllable of the second hemistich (Elwert, 1965: §87) (*césure enjambante*):

(11) qui de s'a.**mi//e** res.pi.te sa joi.(e)
(*Chanson de Berne*, XXIII, l. 6; Elwert, 1965: §87)

Finally, word-final schwa may occur at the cæsura (lyric cæsura) (Elwert, 1965: §88):

(12) que la **ter.re** // de main.te co.lour gai.(e) (*Behaingne*, l. 2)

In this case, the prosody of the line is changed, with the stressed syllable in the third rather than the fourth position in the line. From the point of view of the scansion, only the epic cæsura affects the syllable count of the line: this is the only instance in which a word-final schwa before a word beginning with a consonant is regularly not counted.

Following the standard versification rules of MedFr, the syllable structure given in the syllabic tag could be mapped on to the metrical template using the following algorithm:

Scansion algorithm (provisional)

Map each syllable from the syllabic tag onto a metrical position in the line **except if**:

(a) it is a word-final elidable schwa followed by a vowel;

(b) it is a word-final schwa syllable at the end of the line;

(c) it is a word-final schwa syllable falling directly after an epic cæsura;

(d) it is a word-internal schwa in hiatus (`h`), and the text shows elimination of schwa in hiatus;

(e) it is a vowel in hiatus with a following vowel (`dd`), and the text shows a monosyllabic scansion of vowels in hiatus.

Yet this algorithm is too simplistic. Rules (d) and (e) are problematic because the linguistic changes affecting these environments do not seem to affect all lexical items at the same time: for example, as discussed above, a monosyllabic pronunciation of *oïl* is attested in the 13th century, while a monosyllabic pronunciation of *oïr* is rare even in the 15th century. However, even the other rules, although unaffected by linguistic change, are subject to exceptions. A different approach is required.

2.2.2 Exceptions to scansion rules

Even in texts which generally conform well to the conventional scansion rules, schwa elision is not as regular in practice as it seems from the versification manuals. Firstly, in the majority of texts there are cases in which word-final schwa does not elide before a vowel:

(13) que pur e.les grant joi.**ë** unt (*LaisMarie*, l. 250)

With words of more than one syllable, such cases are exceptions to the general rule of elision before a vowel, and are usually marked by modern editors with a diæresis. With monosyllables, however, non-elision is more common:[37]

(14) au plus lon.gue.mant **que** il pot (*Charrete*, l. 62)

Non-elision here is not usually marked by modern editors. Secondly, in some later texts word-final schwa is deleted in contexts where it would normally not be elided:

(15) a.ler nu**(es)** jam.bes en cha.ppin (*Testament*, l. 1043)

In *Testament*, from the mid-15th century, the final schwa syllable of a word of more than one syllable may be elided where it has no onset (i.e. schwa directly

37. Cf. Elwert (1965: §34), who claims that elision of monosyllables is optional.

follows a vowel). However, this rule is not consistent. For example, in other lines of *Testament*, it may be retained:

(16) di.sant que l'a.voy.**e** maul.dic.(te) (*Testament*, l. 1235)

Rather than an exceptional elision rule, this variation perhaps reflects the beginnings of a long-term linguistic change which would lead to the loss of all final schwa by the ModFr period.

The linguistic changes identified by the syllabic annotation (loss of schwa in hiatus and monosyllabic scansion of vowels in hiatus) give rise to substantial variation in versification practice. Regularly syllabic in the 12th and 13th centuries, by the 15th century schwa in hiatus is in the process of being lost from pronunciation:

(17) S'il **(e)**ust au gré du cuer vi.sé (*BelleDame*, l. 139)

(18) Hyn.ce.lin es.l**(e)**u de Pa.ris (*Testament*, l. 1015)

In *Testament*, this loss is regularly reflected in the syllable count. However, in the earlier *BelleDame* text, variation is attested. Different treatments of hiatus schwa are even found within the same line:

(19) Ne me p**e.**üst v**e**oir au tra.vers (*BelleDame*, l. 160)

Through use of the diæresis, the editor suggests that the schwa in *peust* was retained, while the schwa in *veoir* was not. However, all that can be deduced from the syllable count, which in other respects is regular in this text, is that one schwa was pronounced, and one was not. With regard to the scansion of diphthongs, variation is widespread, as discussed in the previous section (cf. also Lote, 1949–96: III, 115–33).

The variability in the treatment of schwa elision, schwa in hiatus and diphthongs across the corpus means that a simple algorithm of the type suggested above is unsuitable. Instead, the algorithm is given a set of rules, but rather than apply them to every line, it varies the rules applied on a line-by-line basis in order that as many lines as possible show a particular number of syllables. The full set of rules is given below:

R1 Elide all elidable word-final syllables in words of more than one syllable when followed by a word beginning with a vowel.

R2 Elide all elidable monosyllables when followed by a word beginning with a vowel.

R3 Delete all schwa syllables in hiatus (`h` syllables).

R4 Count all vowels in hiatus (`dd`) as one syllable.

Text	Total lines	General rules (lines)	Specific rules (lines)	Unscanned
Charrete	501	R1 (501)		0
MirNDCoin	496	R1 & R2 (494)		2
PassJongl	540	R1 (535)	Not R1 (3); Add R2 (1); Add R3 (1)	0
RoseMeun	504	R1 (501)	Add R2 (3)	0
ComteAnjou	509	R1 (508)	Not R1 (1)	0
LibFortune	496	R1 & R2 (447)	Not R1 (10); Not R2 (8)	31
VoirDit	496	R1 & R2 (493)	Not R1 (1); Not R2 (1)	1
MutFortune	516	R1 & R3 (490)	Not R1 (1); Not R3 (10); Add R2 (6); Add R4 (1)	8
BelleDame	504	R1 & R2 (496)	Not R1 (1); Not R2 (3); Add R3 (2); Add R4 (1)	1
Testament	504	R1 & R3 (498)	Not R1 (1); Add R2 (1); Add R5 (4)	0

Table 4: Summary of scansion for texts in the base subcorpus

R5 Elide all word-final schwa syllables when in hiatus with a preceding vowel (cf. 15).

R1 is common practice throughout the medieval period. R3 and R4 both use the syllabic tagging to take account of linguistic changes; R5 also models a linguistic change. R2 models elision of schwa in monosyllables, a phenomenon variably attested throughout the medieval period.

Before the algorithm is run, the annotator specifies a target line length and cæsura position (if applicable). Depending on the date of the text being annotated, certain rules may be disabled: for example, R3–R5 would normally be unavailable to the algorithm when processing 12$^{\text{th}}$-century texts. The algorithm then carries out a two-stage evaluation. Firstly, it determines which combination of rules yields the greatest number of lines in the text that meet the target line length. Secondly, any lines which do not meet the target length are evaluated individually to attempt to find a combination of rules which does produce the target length.

To illustrate the operation of the scansion algorithm, table 4 provides statistics for the base subcorpus. The column 'General rules (lines)' gives the combination of rules which, when applied to all lines in the text, produces the

greatest number of octosyllabic scansions (number of lines given in parentheses). The column 'Specific rules (lines)' gives the modifications made to the general rules to derive an octosyllabic scansion for the remaining lines. The 'Unscanned' column gives the number of lines for which no octosyllabic scansion can be obtained. For example, applying rules R1 and R3 generally to *Testament* gives an octosyllabic scansion for 498 of 504 lines. One line requires R1 not to apply ('Not R1'):

(20) lui chan.**ge a** u.ne ju.ment (*Testament*, l. 1012)

Table 4 shows that lack of elision of this kind is found in a small number of lines in the majority of texts. One line requires R2 to apply:

(21) pour.veu, se huys y a ne fe.nes.(tre) (*Testament*, l. 1349)

The elision of schwa in monosyllables was acknowledged to be variable. In *MirNDCoin*, *LibFortune*, *VoirDit* and *BelleDame* it is posited as a general rule, which simply means that it applies more often than it fails to apply. With the exception of *MirNDCoin*, *ComteAnjou* and *Charrete*, all texts show variable behaviour of schwa monosyllables. If the rule is general, there are lines in which it fails to apply; if the rule is not general, there are lines where it does apply. The remaining four lines of *Testament* require R5:

(22) a.ler nu**(es)** jam.bes en cha.ppin (*Testament*, l. 1043)

(23) des ans y a de.my**(e)** dou.zai.(ne) (*Testament*, l. 1154)

(24) a me.nu**(es)** gens me.nu**(e)** mo.nnoy.(e) (*Testament*, l. 1651)

(25) que cha.rec.te.ri**(e)** se boit tou.(te) (*Testament*, l. 1686)

The application of R5 is unique to *Testament*, and even here it is not a general rule, since although there are four lines where R5 does apply, there are ten lines to which applying R5 leads to a non-octosyllabic scansion. *Testament* is also the only text to which R3 (loss of schwa in hiatus) applies without exception. In texts from the late 14th century and the first half of the 15th century, R3 application is more variable.

An implicit assumption of the scansion algorithm is that all lines should conform to the target syllable count. By allowing defined areas of variability in the scansion rules, the algorithm attempts to fit as many lines as possible into the target metrical template. In other words, we assume that the text is fundamentally regular, and any apparent irregularities simply reflect variability in the pronunciation of vowels in hiatus and the elision of schwa. For all but one text in the base corpus, this assumption is clearly defensible, since at least 95%

of the lines in the extract conform to general rules.[38] With *LibFortune*, where only 90% of lines conform to general rules and even variability in the scansion rules cannot account for 6.2% of the lines, the judgment is more marginal. The question is whether this is essentially an octosyllabic text, but with much more flexible scansion (in which case, the algorithm can be applied), or whether in fact the lines conform so poorly to the scansion rules that the text is better described as having a different metrical form. With 90% of lines conforming to general rules and 93.8% conforming if some flexibility is allowed, I feel that an octosyllabic analysis is permissible. Metrical annotation is applied to the 93.8% of lines that can be analysed as octosyllables; the remaining lines are left unannotated.

2.2.3 Texts with different metrical norms

While *LibFortune* is considered an octosyllabic text, there are a number of texts in the corpus which do not conform to the general pattern of the verse form. In the case of *Brendan*, the metrical exception is regular: line-final feminine syllables are counted as syllabic:

(26) Don.na Aa.liz, la re.ï.**ne**,
par qui val.drat lei di.vi.**ne** (*Brendan*, ll. 1–2)

A regular exception of this kind is unproblematic for metrical annotation, as the algorithm can easily be modified to count syllables slightly differently in this text. However, since such lines do not conform to conventional octosyllabic meter, they must be tagged explicitly in order to be treated separately or discounted from later analyses.

More difficult to treat are texts which do not have a regular syllable count. This is most common in the Anglo-Norman tradition, and has led to speculation that Anglo-Norman poets used a different meter, perhaps based on counting stressed syllables (see, for instance, Johnston, 1980, 1983). In order to give an idea of the extent of the problem, table 5 shows the percentage of octosyllabic lines in texts from the Anglo-Norman subcorpus.[39] Clearly, 13$^{\text{th}}$-century Anglo-Norman texts do not follow continental versification patterns. This is not to say that the verse is 'faulty' (as argued, for example, by Vising,

38. Editorial practice could contribute to this apparent regularity, although as far as I am aware none of the editions used make systematic metrical 'corrections'. However, editors do not always clearly state the reasons for their interventions, and it is not always clear from variants whether the motivation for an intervention is metrical or not.

39. I.e. with application of R1, R2 permitted; R3–R5 do not apply in continental texts in the 13$^{\text{th}}$ century.

Extract	Total lines	Octosyllables	%
Prothes	504	483	95.8
VieEdmund	499	415	83.1
GuiWarewic	500	251	50.2
Chivalier	504	259	51.4
VieRichard	516	172	33.3

Table 5: Proportion of octosyllabic lines in the Anglo-Norman subcorpus

1923), any more than 12$^{\text{th}}$-century continental verse is 'faulty' iambic tetrameter (the percentage of conformity is similar). It is clear that the meter of Anglo-Norman is different; the rules of the meter, however, are not known.

From the point of view of the present study, there are two possible approaches to these texts. One approach is to exclude them from studies requiring metrical annotation because they do not follow the same meter as their continental counterparts, and hence are not comparable. Alternatively, we could take only those lines which happen to follow continental norms (and hence are comparable) in the hope of shedding some more light on the unknown versification rule of Anglo-Norman. In particular, if the comparable Anglo-Norman lines show stronger rhythmic organization than their continental counterparts, this would lend support to the view that Anglo-Norman developed a more stress-based versification system. I have elected to adopt the second approach. Only lines of Anglo-Norman verse which conform to continental norms are metrically annotated, and only these lines are used in the studies in chapter three. However, the more limited nature of the data from the Anglo-Norman subcorpus means that only tentative conclusions may be drawn, and a more rigorous study must be left to future research.

2.3 Part-of-speech annotation

The third layer of annotation involves labelling the part of speech of each word. In comparison with the structural annotation discussed in the following section, it is relatively straightforward. For the purposes of the present corpus, the part-of-speech layer has two main functions. Firstly, it can be searched in its own right (e.g. a search for `Adj` followed by `N` to locate adjectives followed by nouns). Secondly, it functions as the input to an algorithm which labels higher-level syntactic constituents (cf. §2.5).

2.3.1 Tagset

Three existing MedFr corpora use part-of-speech tagging (NCA, BFM, MCVF). The three corpora all use different annotation schemes. [40] While the annotation schemes are broadly similar, there are some areas of divergence. Firstly, the size of the tagset. Larger tagsets contain more specific tags. The NCA tagset is the most specific, and is alone is specifying morphological detail such as verb person and the case of nouns. The MCVF tagset is generally the least specific. For example, while the BFM and the NCA contain separate tags for personal, adverbial (*i*, *en*), indefinite and possessive pronouns, the MCVF corpus has only one tag for all four types. (All corpora, however, distinguish impersonal pronouns.) However, in some instances, the MCVF tagset is more specific: the BFM tagset, for example, is the only one not to distinguish singular and plural nouns. Secondly, a number of words are consistently classified differently. For example, the MCVF tagset is the only one with a separate tag for quantifiers, such as *tout*, *tant*, *chaque*, *assez*. In the BFM, these are treated either as determiners (when modifying a noun), adverbs (when modifying an adjective) or pronouns (e.g. *du tout*). *Tout* is even treated as a determiner when it precedes another determiner (e.g. *tous les gens*). [41] Finally, there are some words which prove difficult to classify. These 'grey areas' may be found either where a word takes on a different function in a different context (e.g. a nominalized infinitive, or a past participle used as an adjective), or where a word is subject to ongoing grammaticalization (e.g. MedFr negative reinforcers *pas*, *mie*, etc.).

An important consideration when designing the annotation scheme for this study was feasibility. Manual annotation is time-consuming, as the full corpus contains over 250,000 words. Fortunately an automated tagger is available. The cross-linguistic annotation tool TreeTagger (Schmid, 1994) was trained on the manual part-of-speech annotation of the NCA to produce a set of 'definitions' specific to MedFr. [42] The tagger can be used to annotate any MedFr text. While not perfectly accurate, a process of automatic tagging followed by manual

40. For the NCA, scant documentation is available: a list of tags used is available at <`http://julienas.philosophie.uni-stuttgart.de/nca/tagger/tagdoc.txt`> [accessed 21 January 2011] (registration required); see also Stein and Gleßgen (2005) and Stein (2007). For the BFM, the tagset is described by Guillot and Prévost (2010a,b). For the morphological tagset of the MCVF corpus, see Martineau (2009).

41. Guillot and Prévost (2010a: 14) refer to this use of *tout* as a 'prédéterminant'.

42. The parameter files to enable the TreeTagger to annotate MedFr were produced by Achim Stein, and are available at <`http://julienas.philosophie.uni-stuttgart.de/nca/tagger/`> [accessed 21 January 2011] (registration required). For a description of the project that prepared this software, see Stein (2007).

TreeTagger tag	Corpus tag	Description
ADJ	Adj	Adjective
ADV	Adv	Adverb
CON:coord	Crd	Coordinator
DET	Det	Determiner
PRE	Pre	Preposition
PRO	Prn	Pronoun
NOM	N	Noun
PRO	Neg	Negative *ne*
PROCON	Sub	Subordinator
VER	Vfin	Finite verb
VER	Vinf	Infinitive
VER	Vpp	Past participle
VER	Vprs	Present participle

Table 6: Part-of-speech tags: corpus tag vs. TreeTagger

correction proved by far the most rapid means of annotating texts in the corpus. While based on the tagset of the NCA, the annotation produced by the TreeTagger is far less detailed, containing no morphological information (such as person, number, tense or case).

As more detailed part-of-speech annotation is not necessary, the tagset adopted in the corpus differs minimally from that produced by the TreeTagger. The only major difference is in the verbal domain. The TreeTagger tags all verb forms as VER, but I have manually edited these tags to distinguish finite verbs from infinitives and past and present participles. The tagset is shown in table 6.

Where a single orthographic word is a combination of two separate functional forms, the form is given two tags separated by a + sign. For example *du* is Pre+Det, while enclitic pronouns are indicated by a +Prn tag appended to the tag of the host word. Compound determiner and past participle forms (e.g. *ledit*) are also treated in this way.[43] The possibility to double (or even triple) tag ensures that irregular word division does not affect the annotation.

43. As double tagging is usually restricted to separate functional forms, it could be argued this is a stage too far: *ledit* is perhaps best considered to be a single grammaticalized determiner, in the same way that *dedans* and *defors* are tagged simply as adverbs or prepositions. All corpora use some form of Pre+Det tag, but otherwise double tagging is not used.

2.3.2 Treatment of 'grey areas'

In treating 'grey areas', I have generally favoured adopting a single tag for a single form. In this sense, the tagging tends to be morphological rather than morpho-syntactic. The advantage of this approach is that morphological criteria are often more clear-cut than morpho-syntactic criteria, which must also take account of syntactic function. For example, past participles are consistently tagged as `Vpp`, regardless of whether they are verbal or adjectival.[44] In practice, the distinction between adjectival and verbal function can be extremely subjective, as in the following example from *GormIsem*:

(27) mut l'ad nafré al flanc senestre
que tute **est muillee** la suzcele (*GormIsem*, ll. 252–53)

It is not clear how valid it is here to distinguish a stative, adjectival reading (e.g. 'is soaked') from a passive verbal reading (e.g. 'has got soaked'). Equally, nominalized infinitives are consistently tagged as `Vinf`.[45] A similar morphological criterion is applied to words which may introduce subordinate clauses: for example *que*, *come*, *quant* 'when' and *dont* (relative). These are consistently tagged as `Sub`. The post-verbal negative marker *rien* is tagged as `N`, as it continues to function as the direct object of a verb.

However, there are some cases in which the syntactic functions of two forms were considered so different as to justify separate tags. In some cases, this is simply a case of disambiguating homonyms with separate etymological origins (e.g. *ou* 'where' as `Sub` vs. *ou* 'or' as `Crd` vs. *ou* (*en* + *le*) as `Pre+Det`). Demonstrative and indefinite pronouns were consistently distinguished from demonstrative and indefinite determiners, thus words such as *cil*, *cist*, *nul* and *tout* are tagged as `Det` (when modifying a noun) and as `Prn` (when pronominal).[46] *Tout* may additionally be tagged as `Adv` when modifying an adjective or a verb. *Mais* and *ne* were both tagged as `Crd` when functioning as coordinators. Like the MCVF corpus, but unlike the BFM and the NCA, a consistent distinction was made between adverbial and prepositional uses of locative words such as *devant*, *(de)sus* and *(de)fors*. These were tagged as `Pre` when modify-

44. The BFM initially took the same position: 'Dès qu'il existe un infinitif, on étiquette "vpp" [past participle]' (Prévost, 2001: 8).

45. An exception was made for the infinitives *avoir* 'possessions', *savoir* 'knowledge', *voloir* 'will', *pouvoir* 'power' and *devoir* 'duty'. Not only are these common nominalized forms, it was felt that there was significant semantic difference between the verb and the nominalized form. However, with retrospect, this is the type of fine semantic judgment that a morphological approach to part-of-speech tagging is supposed to avoid, and an equally valid case could have been made for leaving these as `Vinf`.

46. This analysis is also found in the BFM (Guillot and Prévost, 2010a).

ing a following noun phrase and elsewhere as `Adv`.[47] In the case of the negative particles *pas* and *mie*, it was felt that linguistic change had produced forms whose meaning and function was sufficiently distinct from the nouns *pas* 'step' and *mie* 'crumb' that a separate `Adv` tag was justified. More difficult is the case of *bien* and *mal*: these were tagged as `Adv` except when clearly nominalized (e.g. by a preceding determiner).

The tagging scheme adopted is sufficient for its purpose, the provision of basic categorial information both to the corpus user and to the algorithm which labels higher level syntactic constituents (see below, §2.5). However, the tension between the morphological principle of 'one form one tag' and a desire to use separate tags in cases where one form clearly has more than one function creates an annotation system which, while consistent in its application, is not consistent in its principles.[48]

2.3.3 Implementation

For most extracts, part-of-speech tagging was carried out by the TreeTagger, with the tags subsequently converted to the corpus tagset using a simple script. Disambiguation of the `VER` tag was carried out manually. The tags were then checked manually to correct the many errors in the automatic tagging.

Twenty-five of the extracts in the corpus are also found in the NCA with manual part-of-speech tagging already completed. Two further extracts (*MemCommyn* and *ChRoland*) were available manually tagged using the BFM system. This made the part-of-speech tagging task a simple matter of tag conversion followed by a small number of manual changes required by the difference in design of the tagsets (in particular, treatment of subordinators, the `Neg` tag and the `Pre` vs. `Adv` distinction).

2.4 Structural annotation: previous models

Structural annotation is the most difficult aspect of the annotation of any corpus to design and implement. Unlike syllabic and part-of-speech annotation which simply require a tag to be placed on each word in the text, structural annotation requires the text to be organized into a syntactic tree, which is

47. This distinction is vital for correct automated identification of prepositional phrases; see §2.5.

48. The recently reopened BFM adopts both a morphological and a morpho-syntactic part-of-speech tag (Guillot and Prévost, 2010a: 3). The release of the documentation postdates the design of the present corpus, but I would consider introducing such a scheme at a later date.

then annotated. Unfortunately, unlike part-of-speech tagging which takes as its base the relatively common ground of traditional grammatical categories, there is less common ground found in approaches to syntax. As a consequence, different corpora will take very different approaches to syntactic annotation.

Structural annotation in the present corpus is required for several purposes. Firstly, it is needed for studies of the syntactic structure of verse in chapter three, in particular with regard to phonological phrasing. Secondly, it is needed to identify particular word orders for the studies in chapter four. Finally, it is used as a basis for creating 'pseudo-verse' out of prose texts, a procedure documented in chapter three (cf. Gasparov, 1987; Noyer, 2002) All tasks will require searching and analysis of the syntactic structure of the corpus texts, therefore, a treebank approach is necessary.

Building a syntactically annotated corpus is extremely resource-intensive. In this section, I review a number of approaches to the problem, all of which were developed by major research projects. For the present thesis, a more minimal annotation scheme is required. However, the design draws extensively on techniques adopted in other corpora.

2.4.1 Penn system

The most influential treebank format used in the Americas and the UK is the Penn Treebank format. In addition to its use with corpora of contemporary and historical English, the format is used in the only currently available treebank containing MedFr texts, the MCVF corpus (Martineau et al., 2010).[49] The Penn treebank system is based on generative analyses of syntax current in the early 1990s. A crucial aspect of the linguistic theory underpinning these approaches is that hierarchical structure determines linear word order. As a consequence, the annotation organizes the words in the text into a hierarchical structure which reflects the surface word order.

Below are two sample sentences from Penn treebanks, shown both in their physical form as bracketed text and (partially) in tree form:

(28) (a) Which papers did you file without reading?

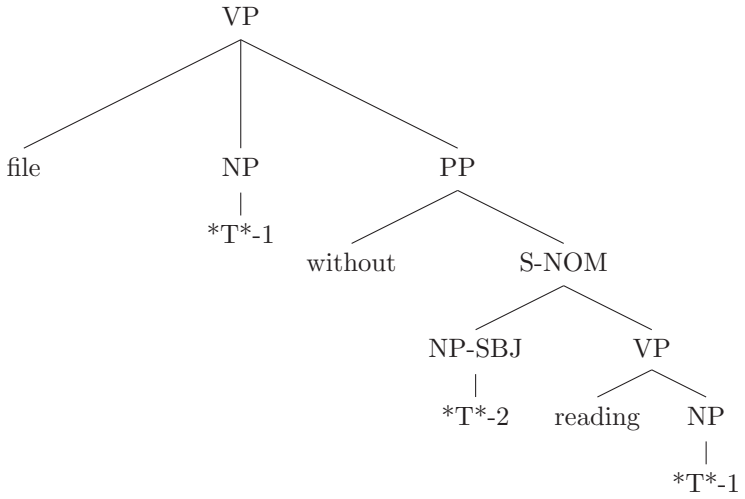
(b) `(SBARQ (WHNP-1 Which papers)`

49. The Penn Treebank project (<`http://www.cis.upenn.edu/~treebank/`> [accessed 21 January 2011]) produced a number of annotated corpora of modern English, including a syntactically annotated version of the Brown corpus (cf. Kučera and Francis, 1967). Historical corpora of English based on the Penn system include the *Penn-Helsinki Parsed Corpus of Middle English* (Kroch and Taylor, 2000), the *York-Toronto-Helsinki Parsed Corpus of Old English Prose* (Taylor et al., 2003) and the *Parsed Corpus of Early English Correspondance* (Taylor et al., 2006).

```
(SQ did
    (NP-SBJ-2 you)
    (VP file
        (NP *T*-1)
        (PP without
            (S-NOM (NP-SBJ *-2)
                   (VP reading (NP *T*-1))))))
?)
```

(Penn Treebank; Bies et al., 1995: 68)

(c)

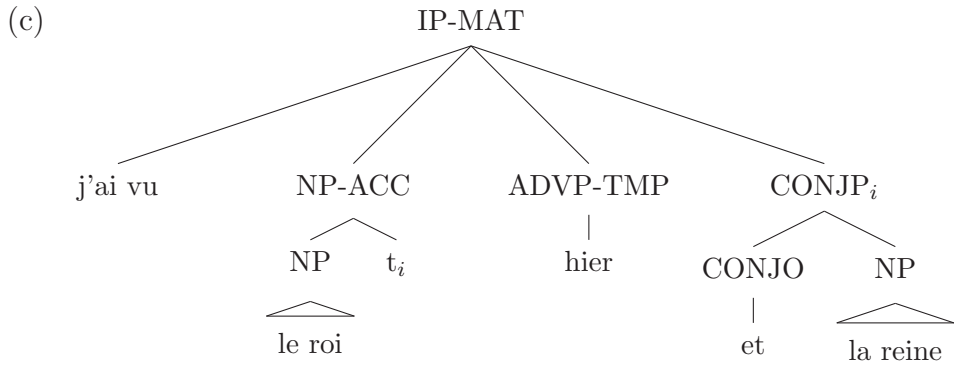


(29) (a) J'ai vu le roi hier et la reine.

(b)
```
( (IP-MAT (NP-SBJ J')
          (AJ ai)
          (VPP vu)
          (NP-ACC (NP le roi)
          (CONJP *ICH*-1))
          (ADVP-TMP hier)
          (CONJP-1 (CONJO et)
                   (NP la reine))))
```

Syntactic annotation from MCVF corpus [50]

50. <http://gtrc.voies.uottawa.ca/manuel/syntax-manual-fr/index.htm> [accessed 21 January 2011], section 'Conjoints extraposés'.



Since the underlying analysis is transformational, the Penn Treebank format makes extensive use of traces and other null constituents. For example, in (28), *which papers* is linked to two movement traces which are the complement of the verbs *file* and *reading*. These traces show that *which papers* is underlyingly the direct object of these verbs, although it does not occupy postverbal position in the surface word order. Indeed, in (28), there is no explicit 'object' tag: the object constituent is indicated only by the structure of the tree. Note also the null subject of *reading*. In the more recent MCVF tagging scheme in (29), functional annotation is more widespread, and the direct object is explicitly marked with the suffix `-ACC`. The structure is also much flatter: for example, the `VP` (verb phrase) tag is no longer used. However, here too null constituents are used: the discontinuous coordination *le roi [...] et la reine* is unified by the use of an `*ICH*` ('interpret constituent here') tag. Respect for linear word order in the hierarchical tree structure makes such phonetically null traces virtually inescapable. Particularly in the original Penn scheme, the annotation introduces structure not visible in the text: in (28), there are seven words, but eleven constituents. While powerful, this scheme is not the 'minimal' solution required.

2.4.2 Paris 7 Corpus

The largest treebank of contemporary French was built at Paris 7 by a team led by Anne Abeillé.[51] While a transformational analysis is not used, the tree structure still follows the linear word order of the text. Unlike the Penn system, neither null constituents nor traces are used. Functional relations such as subject and object are explicitly tagged.

Below I reproduce a sample from the Paris 7 treebank, both in physical form (an XML tree) and partially in tree form:

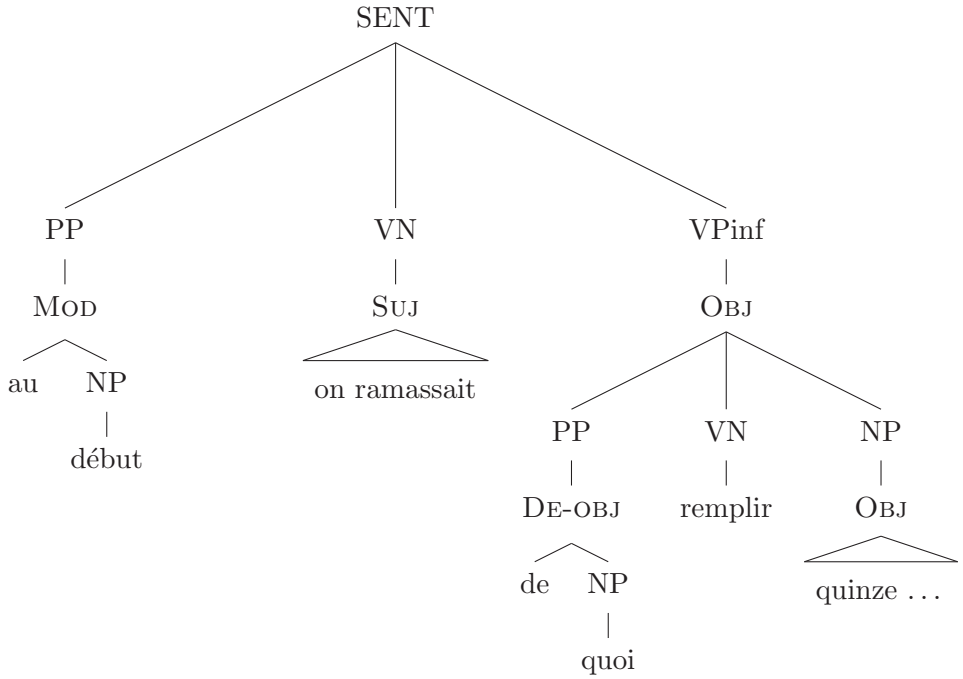
51. Homepage: <`http://www.llf.cnrs.fr/Gens/Abeille/French-Treebank-fr.php`> [accessed 21 January 2011]. See Abeillé et al. (2003) for more information on its construction.

(30) (a) Au début, on ramassait de quoi remplir quinze sacs poubelle, indique Roger, ouvrier à la régie.

(b)
```
<SENT>
  <PP fct="MOD" >Au <NP>début</NP></PP>,
  <VN fct="OBJ" [sic.]>on ramassait</VN>
  <VPinf fct="OBJ">
    <PP fct="DE-OBJ">de <NP>quoi</NP></PP>
    <VN>remplir</VN>
    <NP fct="OBJ">quinze sacs_poubelle</NP>
  </VPinf>,
  <Sint>
    <VN>indique</VN>
    <NP fct="SUJ">Roger, <NP>ouvrier <PP>à
        <NP>la régie</NP></PP></NP></NP>
  </Sint>.
</SENT>
```

Paris 7 treebank sample parse. [52]

(c)



A variety of syntactic functions are tagged (e.g. `SUJ` (subject), [53] `OBJ` (object),

52. Source: <`http://www.llf.cnrs.fr/Gens/Abeille/example2.txt`> [accessed 21 January 2011]

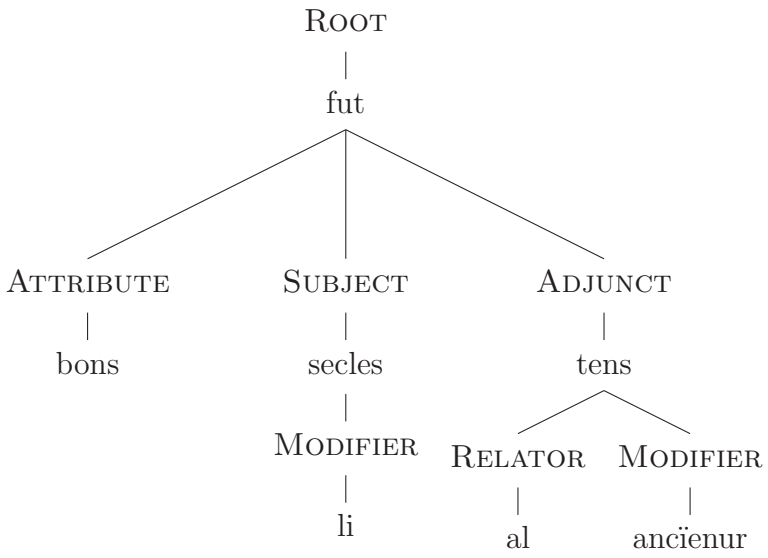
53. The example provided by the corpus website uses the tag `OBJ` for the constituent *on ramassait*, which in my view is erroneous. For clarification, a feature of the corpus design is that subject and object clitics are grouped with the verb, and the function tag of the VN

`DE-OBJ` (complement introduced by *de*)). Unlike in (28), no attempt is made to associate functions with structural positions, thus avoiding the need for movement traces. Constituent type tags are descriptive: the tag `VPinf`, for example, contains basic morphological information. There is, however, one major drawback: the lack of any null constituent of the `*ICH*` type means that discontinuous relations cannot be marked. For a highly configurational language such as ModFr this seems to have been unproblematic, but the same cannot be said for MedFr.

2.4.3 Dependential analyses

Dependential treebanks are favoured in Europe, and have been built for a variety of languages, most influentially Czech (the Prague dependency treebank). In a dependential approach, the tree structure is based not on the linear word order, but on the syntactic relations between words in a clause. Linear word order must therefore be encoded separately. For example, the clause *bons fut li secles al tens ancïenur* (*Alexis*, l. 1) could be analysed as follows:

(31)



The notion of 'dependency' defines the hierarchy: for example, the form of the preposition and determiner form *al* and the adjective *ancïenur* depends on the gender of the noun *tens*, and so is treated as a daughter node. The finite verb defines clausal relations such as subject and object and thus is the root of the clause. Since these dependency relations are fundamental to the analysis, it is problematic if they do not exist. For example, if all elements of a clause are assumed to depend on the finite verb, it becomes impossible to posit structure

group reflects the function of the clitics contained within it, not the function of the group as a whole.

Corpus	Structure	Traces	Functions	XP tags
Penn	Single tree	yes	optional tag	yes
Paris 7	Single tree	no	optional tag	yes
Dependential	Tree and linear	no	structural	no

Table 7: Comparison of treebank features

in verbless clauses, particularly where null elements are excluded.

Work is currently in progress on a dependential corpus of MedFr: the *Syntactic Reference Corpus of Medieval French*, under the direction of Sophie Prévost and Achim Stein. [54] No documentation has been released detailing the annotation scheme at time of writing.

Summary

Table 7 summarizes the key features of the tagging schemes discussed. Differences are found in (i) the format of the data stored (i.e. whether or not the tree respects linear word order), (ii) the use of traces or null constituents (iii) the marking of the syntactic function of constituents (e.g. subject, object, adjunct, locative) and (iv) whether constituents are tagged for their syntactic type (e.g. noun phrase, prepositional phrase, verb phrase; column 'XP tags').

2.5 Structural annotation: current corpus

The annotation scheme adopted in the corpus has the following key structural features:

(i) Tree independent of linear word order
(ii) Hierarchy based on constituency
(iii) Functional annotations explicitly marked
(iv) No traces or null elements
(v) Tree structure built by annotator
(vi) Tree nodes labelled automatically from part-of-speech tags.

With regard to point (i), I follow dependential models. The rigidity of a tree respecting linear word order proves difficult for annotating discontinuous constituents (see 29 above), of which there are many in MedFr. Traces and null

54. Project summary at <https://listes.cru.fr/wiki/srcmf/> [accessed 21 January 2011]

elements are then no longer necessary (iv), in keeping with the goal of producing a minimal scheme of structural annotation. However, unlike dependential models, the tree hierarchy is based on constituency (ii), in common with transformational models and in particular the Paris 7 treebank. Functions such as subject and object are marked explicitly (iii), rather than structurally encoded (as in 28).

While the tree structure must be built in full by the annotator (v), the scheme is in other respects minimal. The number of tags the annotator must add to the tree is very small. Only basic functional tags are used (e.g. subject, complement), and the function of most constituents is left unmarked. This annotation can of course be enriched at a later stage. More innovatively, tags showing the syntactic type of the constituent (e.g. noun phrase, clause) are automatically generated from the part-of-speech tagging (vi), and do not have to be marked explicitly. Finally, not all the texts were annotated. Priority was given to the annotation of the early-8 and base subcorpora in order to focus on the development of octosyllabic narrative verse. [55] The prose subcorpus was also structually annotated so that the studies in chapter four could compare verse and prose texts. In total, 27 of the 87 extracts were structurally annotated with manually corrected part-of-speech annotation. The remaining texts were fully annotated syllabically and metrically, and were automatically part-of-speech tagged using the TreeTagger.

I begin by defining the word and the sentence, the maximal and minimal unit of the analysis. I then illustrate the tagging scheme and discuss its core features through the annotation of a complex worked example.

2.5.1 Words and sentences

The minimal unit of syntactic annotation is the word. The maximal unit is the sentence. Neither the word division nor the punctuation of modern critical editions reflects the word division or punctuation of the original manuscript, and both follow arbitrary and anachronistic conventions (for example, the use of apostrophe to mark elision of schwa in articles or object pronouns before a vowel). The corpus linguist is faced with a choice: ignore the editor's punctuation and word division and offer principled linguistic reasons for an alternative division into words and sentences, adopt the editor's punctuation and develop a linguistic analysis capable of dealing with inconsistency, or follow

55. Moreover, as the studies in chapter three were carried out, it became apparent that the most interesting findings would emerge from these texts.

the manuscript exactly.

Of the other MedFr corpora, the NCA and the MCVF corpus both elect to standardize word division across the source texts, even where this involves splitting units which are consistently written as single words by modern editors and frequently in manuscript sources too (e.g. *jol* (*jo+le*) becomes *jo l* in the NCA). Unlike the NCA, the MCVF corpus marks explicitly where a word has been split (Martineau, 2009: 'Découpage et unification de mots'). This has the advantage that a 'one form one tag' approach to part-of-speech annotation can be adopted. However, it also assumes a linear view of morphology, and the decision of the MCVF annotators to split forms such as *au* into *a* (preposition) and *u* (determiner) does not respect the morphological structure of what is, at least since the vocalization of *l* ceased to be an active rule, a single complex morpheme. The solution adopted in the present corpus, as discussed in section 2.3.1, is to apply a double functional tag to a single form. There seems no reason therefore to alter the word divisions favoured by modern editors.

For division into sentences, both the MCVF corpus and the *Syntactic Reference Corpus of Medieval French* project have developed their own conventions.[56] Particularly difficult for a principled treatment of sentence division is the treatment of coordination, and the attachment of clauses introduced by *que* or *car*. In my own corpus, I have elected to follow the editor's division of the text into sentences: these are units delimited by a full stop, colon or by an exclamation or question mark followed by a capital letter. Combined with the use of editorial word division, the main advantage of this approach is that the annotated text follows a fixed, published standard. Furthermore, it is not straightforward to offer clear linguistic principles governing sentence division, and different corpora choose different solutions. For example, the Penn system adopted by the MCVF corpus distinguishes the coordination of verbs with arguments (e.g. 'he sang and danced the polka'), which are treated as separate sentence tokens, from the coordination of intransitive verbs (e.g. 'he came, saw and conquered'), which is treated as a single token (Santorini, 2010: 'Syntactic annotation: division into sentence tokens'). The *Syntactic Reference Corpus of Medieval French* scheme, on the other hand, does not at present permit any coordination of finite main clause verbs. Moreover, once split into separate tokens, the clauses are analysed separately in any corpus search. Following the editorial reading of a sentence, which almost invariably gives rise to longer

56. For a basic treatment of the analysis adopted by MCVF, see Santorini (2010: 'Syntactic annotation: division into sentence tokens').

base units than a linguistic approach, allows a more flexible treatment of inter-clausal relations. Below, I illustrate how main clauses within a single sentence can be annotated either as conjoined with or as 'continuations' of the previous clause.

With the definition of words and sentences clarified, annotation of the following complex sentence will be discussed.

(32) L'en ne doit pas croire fol home
de la value d'une pome;
blasmer le doit l'en et reprendre
ainz qu'il li laist folie emprendre;
et je fui fox, et el me crut,
onques par lui mes biens ne crut.

(*RoseMeun*, ll. 4109–14)

2.5.2 Marking constituents within clauses

Syntactic annotation is based on groups. In the scheme adopted for this corpus, groups are made up of lexical heads and their modifiers. Consider the following extracts from the worked example:

(33) `[ fol home ]`

(34) `[ par lui ]`

(35) `[ mes biens ]`

(36) `[ et je fui fox ]`

Examples (33–36) illustrate sample basic groups. In (33) and (35), a noun is grouped with agreeing adjectives and determiners; in (34), a pronoun is grouped with its associated preposition, while in (36) a finite verb is grouped with the other arguments of the clause. The grouping is uncontroversial and straightforward.

In fact, the groupings in (33–36) are so straightforward that the correct tag for the group trivially follows from the part-of-speech tags of the contained words. Examples (33) and (35) contain nouns and are noun phrases. Example (34) contains a nominal (a pronoun) and a preposition, and so is a prepositional phrase. Example (36) contains a finite verb and is a clause. These implications form a phrase structure rule in reverse:

`Adj`, `N` —> noun phrase (e.g. 33 above)
`Crd`, `Prn`, `Vfin`, `N` —> main clause (e.g. 36 above)

Note also that if a group is made up of a head and its modifiers, the order of the elements in the group is not important in determining the group type. For example, in (33), it is irrelevant for the group type whether the noun follows the adjective or the adjective follows the noun; in (36), the group would still be a clause even if the word order of the constituents were different. A simple algorithm was devised to identify both the lexical head of a group (noun, verb, adjective or adverb) and the presence of any prepositions or subordinators, and thus generate the correct tag for the group (e.g. noun phrase, prepositional phrase, finite clause, subordinate clause, etc.).

Below is the worked example divided into clauses with all the clausal constituents marked:

(a) `[ L en ] ne doit pas [ croire [ fol home ]`
`[ de la value [ d une pome ]]]`
(b) `blasmer le doit [ l en ] et reprendre`
(c) `ainz qu il li laist [ folie emprendre ]`
(d) `et je fui fox`
(e) `et el me crut`
(f) `onques [ par lui ] [ mes biens ] ne crut.`

The grouping rules adopted require the splitting of complex nominals such as *de la value d une pome* in (a) into two groups, the second embedded in the first. Failure to split the group would create a group with two heads, *value* and *pome*, but modifiers that apply only to one of them (*de la* and *d une*). Infinitives in (a, c) are grouped together with their complements, and infinitives with complements are treated as a single group at clausal level (i.e. as an infinitival clause).

Although the tag denoting the syntactic type of a group is automatically generated, functional tags must be supplied by the annotator. A functional tag denotes the function of a word (or group) within the containing group. These are mainly restricted to the core verbal arguments subject (`@s`), complement (`@c`) and attribute (`@a`) (either of subject or of complement). Having marked the clausal constituents, functional tags may now be added:

(a) `[@s L en ] ne doit pas [@c croire [@c fol home ]`
`[ de la value [ d une pome ]]]`
(b) `blasmer@c le@c doit [@s l en ] et reprendre`
(c) `ainz qu il@s li laist [@c folie@c emprendre ]`
(d) `et je@s fui fox@a`
(e) `et el@s me crut`
(f) `onques [ par lui ] [@s mes biens ] ne crut.`

Notationally, a functional tag is preceded by the `@` symbol, and is placed after the opening bracket of a group or after a single word. For example, `[@s L en]` in (a, b) denotes that the group *l'en* is the subject of the clause, just as `je@s` in (d) denotes that *je* is the subject. The complement tag is used for direct objects, as in (b, c),[57] for direct infinitival complements (a, b, c), and for core prepositional complements of intransitive verbs (e.g. directional complements of motion verbs). The attribute tag is used mainly for arguments of the copula, as in (d). This annotation is not intended to be comprehensive as a sophisticated system of functional tagging is not necessary for the corpus studies envisaged.

2.5.3 Marking clauses within sentences

Since the corpus adopts the editor's sentence division, a sentence may consist of multiple clauses. The corpus models three types of interclausal relation:

- subordination
- coordination
- continuation

A clear example of subordination is found in the worked example:

(37) `[ blasmer le doit l en et reprendre`
`[ ainz [ qu il li laist folie emprendre ]]]`

The subordinate clause is marked as a group, including the subordinator *qu*, and is embedded within the main clause. As the group contains a finite verb and a subordinator it is automatically identified as a subordinate clause.[58]

57. In (b), the clitic *le* is the direct object of the infinitive *blasmer*. Clitic climbed arguments such as this form a standard exception to the rule that infinitives should be grouped with their arguments. In corpus searches, such structures are nevertheless easily identifiable: where both an infinitive and an object clitic are marked as `@c` in a main clause, the clitic is always a complement of the infinitive.

58. There are a small number of clauses which are functionally subordinate, but have no subordinator:

(i) n'i at celui **paor n'ait grant**

(*PassJongl*, l. 291)

The clause *paor n'ait grant* is a relative clause, containing a subjunctive verb but unmarked by a subordinator. Here, the annotator must explicitly mark subordination using a function tag `@sub`. Secondly, there are rather more cases where *que* does not appear to function as a subordinator:

(ii) mains en deceit par sa promesse
qu'el promest tel chose souvent
dom el ne tendra ja couvent

(*RoseMeun*, ll. 4044–45)

In the clause *qu'el promest...*, *qu'* functions as a connective rather than a subordinator. As *qu'* is consistently part-of-speech tagged as `Sub`, these clauses are consistently tagged as

Note that the adverb *ainz* is split from the subordinator. The subordinator *qu* is considered a constituent of the embedded clause, but *ainz* cannot be: since groups are unordered, it would be indistinguishable from clause-internal adverbs.

For the remaining clausal links, the annotator may make use of two relations: coordination (`@&`) and continuation (`@+`). Coordination indicates a measure of syntactic unity between clauses. For example, where the coordinator is present and there is subject continuity, there is a possibility that the second conjunct contains a subject gap:

(38) Entr'ex en demeinent grant plet
`[@&` et demandent qui ce a fet `]` (*Thebes*, ll. 265–66)

This example from a 12th-century text shows subject continuity, presence of the coordinator and a verb-first second conjunct. The lack of an initial constituent in the second conjunct, contrary to the usual verb-second structure of main clauses in the 12th century, suggests that the second conjunct is not a syntactically independent clause. Equally, coordination can be used in the case of 'listed' clauses, with a similar syntactic structure:

(39) ne set s'il est `[@&` ou s'il n'est mie, `]`
`[@&` ne ne li manbre de son non, `]`
`[@&` ne set s'il est armez ou non `]` (*Charrete*, ll. 716–18)

Coordination is also used where the clauses are contrastive or contradictory:

(40) Phenenna out enfanz plusurs, `[@&` mais Anna n'en out nul. `]`
(*QuatrLivre*, 1:2)

Finally, coordination can be marked in cases where there is clear propositional unity between two clauses:

(41) commandés `[@&` et je le ferai `]` (*VoirDit*, l. 352)

In other cases, a continuation relation is marked (notationally, `@+`). Each clause may have only a single continuation, it must begin at the end of the clause and continue to the end of the sentence. Essentially, it is one step down from marking a full sentence boundary. Yet marking a full sentence boundary is undesirable, as it would separate clauses which are linked at a looser, semantic level. Common tokens of continuation are explanatory *car* clauses.

subordinate. Varying the part-of-speech tag on *que* is undesirable without clear diagnostics to distinguish its various functions. The solution is to annotate such clauses as a continuation of the main clause much as would be done with a clause introduced by *car*. In corpus searches, the continuation tag distinguishes these clauses from other subordinates.

These are not coordinated and rarely show subject continuity, but are strongly semantically linked to what precedes:

(42) Ne me tenez pur fille Belial, `[@+` kar sobre sui `]` (*QuatrLivre*, 1:16)

In this case, the *kar* clause explains or justifies the main clause.

It could be objected that the distinction drawn between coordination in (41) and continuation in (42) is a fine one. I defend in principle the need to use two distinct inter-clausal relations below the level of the sentence: cases of possible subject gaps (38) and those of clear syntactic parallelism (39, 40) should be kept distinct from clauses between which the link is purely semantic, as in (42). However, a lack of clarity in defining the difference between the two relations proved problematic for annotation of the corpus. For example, the clause *et el me crut* in the worked example is tagged in the corpus as a coordination, on the grounds of high propositional unity:

(43) `[ et je fui fox [@& et el me crut ]]`

Yet this is a rather subjective decision, depending more on how the annotator understands 'high propositional unity' than any more objective criteria. In future revisions, I would perhaps look to restrict the coordination relation *only* to a clearly defined subset of cases in which clauses are syntactically linked (possible subject gaps, structural parallels), with the remainder of relations tagged as continuations. This remains an area for future development and research. However, as the I do not focus in depth on inter-clausal relations, difficulties with the annotation in this area do not pose a problem for the present thesis.

Returning to the implementation of the two relations, the second conjunct of a coordinated structure is embedded into the first. Embedding rather than juxtaposition must be used following the principle that groups are treated as unordered sets.[59] Thus, if two conjuncts are juxtaposed, this implies that their ordering is irrelevant. Yet this is not the case: the second conjunct may be impoverished compared to the first by processes such as gapping or elision, as in the following example:

(44) [. . .] ben seiez purpensez
De colps ferir, **de receivre et de duner**
(*ChRoland*, ll. 1177–78)

Here, the second and third conjuncts *de receivre* and *de duner* omit the object *colps*, present in the first conjunct.

59. An embedded analysis of coordination is also used in the MCVF corpus (Santorini, 2010: 'Syntactic annotation: Conjunction').

Aside from the problematic case of *et el me crut*, the remainder of clausal relations in the worked example are continuations. All clausal relations are marked below:

(45)
```
[ L en ne doit pas croire fol home
de la value d une pome
[@+ blasmer le doit l en et reprendre
[ ainz [ qu il li laist folie emprendre ]]
[@+ et je fui fox [@& et el me crut ]
[@+ onques par lui mes biens ne crut. ]]]]
```

2.5.4 Discontinuity

An advantage of separating structural annotation and linear word order is that groups may be discontinuous. This is necessary to show the coordination of the infinitives *blasmer* and *reprendre* in the worked example:

(46)
```
le doit l en [ < blasmer > [@& et reprendre ]]
```

Marking a discontinuous structure is notationally problematic: in (46), I have 'reunited' the discontinuous coordination using angle brackets to show the moved text. Discontinuous groups are formed for any number of phenomena (floating quantifiers, disconnected relative clauses, infinitival clauses with a fronted infinitive, etc.) where it is necessary to group a lexical head with its modifiers.

2.5.5 Full worked example

Below is the worked example annotated in full:

(47)
```
[ [@s L en ] ne doit pas [@c croire [@c fol home ]
[ de la value [ d une pome ]]]
[@+ le@c doit [@s l en ] [@c < blasmer > [@& et reprendre ]]
[ ainz [ qu il@s li laist [@c folie@c emprendre ]]]
[@+ et je@s fui fox@c [@& et el@s me crut ]
[@+ onques [ par lui ] [@s mes biens ] ne crut. ]]]]
```

Chapter summary

In this chapter, I have presented the method used to build a corpus of texts suitable for the studies in chapters three and four. The uneven nature of the philological record poses methodological problems for the diachronic

approach adopted in the present thesis, which relies upon the comparison of similar texts across time. In part, these problems were resolved by structuring the corpus into ten subcorpora, each containing texts classifed according to form, dialect and text type. Where the only available texts are problematic, I have highlighted this in the preceding discussion, and these problems will be borne in mind in interpreting the results from the studies in chapters three and four. The proposed studies also require machine-processing of the corpus texts to derive quantitative results, and this has been implemented by means of a detailed multi-layered annotation system. To make it feasible to build such a corpus with available resources, a number of aspects of the annotation were partly automated (part-of-speech annotation, metrical annotation, labelling of syntactic constituents). The present corpus is the only corpus containing French texts with both metrical and syntactic annotation. As a result, the analyses in chapters three and four are able to use quantitative data to study the interaction of syntax and prosody in MedFr in a way that has not previously been possible.

Chapter 3

Evidence from versification

In chapter one, it was suggested that early octosyllabic verse texts, from the 11th century and the first half of the 12th century may have shown a form of syllabo-tonic versification. Syllabo-tonic versification of this kind, I argued, is incompatible with stress deafness, and hence with a group stress system. This chapter has two main aims, to examine both the evidence for syllabo-tonic versification in the earliest texts, and to establish a precise timeline for the disappearance of rhythmic organization from French versification. In addition to considering the organization of word stress, I will also investigate the possibility that the rhythm of early octosyllabic texts is related to a tendency to place a PhPh boundary in the centre of the line. In chapter one, it was shown that most linguistic changes clearly incompatible with lexical stress, such as diphthongization of primary stressed syllables, pre-date the textual record. If it can be shown that the versification of the earliest texts is at least in part dependent on the position of stressed syllables, then this would suggest that stress deafness effects, and hence group stress, had not yet emerged in MedFr. If successful, the timeline established in this chapter can be used to inform the studies in chapter four, which aim both to examine the linguistic conditions causing the emergence of group stress, and the consequences of this change.

The chapter begins with a methodological discussion, which describes how the corpus annotation is used to reconstruct aspects of the stress system in order to analyse how stress is used in versification. Firstly, the prominent vs. non-prominent syllable distinction is presented, a reconstruction which is based on an assumption of word stress. The analysis of the position of prominent syllables in the line is the basis of all studies of rhythmic organization in this chapter. Secondly, a method for reconstructing PhPh edges is presented. Four studies follow. Study 1 examines the rhythmic development of eight-, ten-

and twelve-syllable verse in the MedFr period. Study 2 focuses only on the octosyllable, and considers the effect of dialect and text type on the rhythm of verse. Study 3 examines the view that octosyllabic verse showed a regular mid-line break, similar to the cæsura of ten- and twelve-syllable verse, by studying the position of PhPh edges in the line and the position of the PhPh-final stress. Finally, study 4 compares genuine octosyllabic verse to 'pseudo-verse' sections drawn from prose texts, in order to test whether the patterns of development observed in studies 1 and 3 are unique to verse, or whether they are simply an artefact of wider linguistic change.

1 Methodology

1.1 Labelling prominent syllables

A prominent syllable is defined as having a stronger underlying stress than its neighbours. The rules used to generate the position of prominent syllables are based only on the syllabic and metrical annotation of the corpus. [1] Syntactic annotation is not used. The labelling of prominent and non-prominent syllables ignores any effect that the presence of higher-level prosodic constituents may have had on the realization of stress. Therefore, prominent syllables only reflect the position of word-level stress.

The concept of a 'prominent syllable' is not the same as that of a 'stressed syllable'. Prominent syllables are generated by a clearly defined regular set of rules. Stressed syllables are a historical reality, unrecorded in written texts, and almost certainly subject to varied realizations across speakers and time. The prominent syllable is a consistent reconstruction of the likely position of stress.

The method used to label prominent syllables owes much to previous approaches to the rhythm of MedFr verse, particularly Le Mée (1978), Pensom (1982, 1985, 2000) and Noyer (2002). Labelling is carried out in two stages. The first stage marks the position of etymological primary word stress. Additionally, reduced (schwa) syllables are marked explicitly as unstressed. This gives rise to a three-way typology:

Type 2:

Syllable bearing primary word stress (e.g. *Ma.**ri**.e*, *de.man.**der***, ***roi***).

1. Limited use of part-of-speech annotation assists in the disambiguation of homographs.

Type 1:

Non-schwa syllables which do not bear primary stress (e.g. ***Ma****.ri.e*, *de.****man****.der*). Includes some functional monosyllables assumed not have developed under primary stress.

Type 0:

Unstressed syllables with a reduced vowel (tagged as `s`) (e.g. *Ma.ri.****e***, ***de****.man.der*).

Two aspects of this labelling in particular call for comment. Firstly, I do not label a fixed secondary word stress. In vulgar Latin, this was realized on the initial syllable of a word (Pope, 1952: §216). However, it is unclear that this was still realized in MedFr: we saw in chapter one (§3.1.2) that initial syllables could undergo vowel reduction to schwa. Secondly, a number of monosyllabic function words are deemed 'unstressed', and hence classified as type 1. Previous studies make a similar distinction between stressed and unstressed monosyllables: for Le Mée (1978) and Noyer (2002) it is arbitrary; for Pensom (1982), it is based on a criterion of frequency (i.e. more frequent words are less stressed). In the present study, I have elected to use a criterion of etymological stress with a measure of flexibility. In general, only those monosyllables whose form indicates an unstressed historical development are included:

(a) Determiners / object pronouns: *la li les lu lo los las*

(b) Prepositions (with determiners): *a as aus au al als aux*, *des du dou del*, *en an es ou el*

(c) Weak non-subject pronouns: *lur lor leur lour*[2], *y i*, *en an*, *vus vos vous voz*, *nus* etc.

(d) Possessive determiners: *mes mi mis mon men mun ma*, *tes* etc., *ses* etc.

(e) Conjunction: *et e*

The list contains weak forms of non-subject pronouns (a, c), all definite and possessive determiners (a, b, d), preposition plus determiner forms of *a*, *en* and *de* (b) and the conjunction *et* (e). Note that only forms containing a full vowel are listed; monosyllables containing a schwa vowel are of type 0. Developments characteristic of primary stressed syllables, such as diphthongization, are not attested. Moreover, some determiners and personal pronouns show atypical phonological developments, such as the loss of the initial syllable of

2. The orthography of the *leur, lour* form suggests a historically stressed development. However, most studies analyse *lour, leur* as equivalent to historically unstressed *lor* (e.g. Pope, 1952: §833; Moignet, 1965: 53; Price, 1998: 143). *Lui* on the other hand is clearly a historically stressed form, and is thus not included here.

the etymon ILLE (Pope, 1952: §837) in the determiners and third person object pronouns. The list of type 1 monosyllables is restrictive in comparison to the forms deemed unstressed by previous studies (Le Mée, 1978; Noyer, 2002). In particular, I avoid any criteria reliant on syntactic context. Noyer (2002: 132), for example, considers all pre-verbal subject pronouns to be unstressed 'when in the syllable preceding the verb' (but not when preceding non-subject pronouns), and also auxiliary uses of *avoir* and *estre*. Le Mée (1978: 46–47) goes even further, assuming that stress on words may be 'weakened' if they are not final in the rhythmic group (e.g. finite verbs followed by stressed monosyllables such as *pas* or *bien*). The use of syntactic criteria is not compatible with the purely word-based view of prominent syllables that I have chosen to adopt, and using such criteria risks confusing an underlying lack of word stress (i.e. clitichood) with the non-realization of word stress due to position within the phonological phrase (i.e. group stress).

Stage one labelling applied to three lines of *BelleDame* is given below:

(1)

0	2	1	2	1	2	0	2	(0)
je	vueil	lai-	ssier	aux	au-	tres	fai-	re

(*BelleDame*, l. 35)

(2)

1	2	0	0	2	0	0	2
et	gai-	res	de	bruit	ne	me-	noit

(*BelleDame*, l. 84)

(3)

2	0	2	2	2	2	0	2	(0)
car	de	plus	beaulx	jeux	on	se	la-	sse

(*BelleDame*, l. 382)

Note the use of type 1 monosyllables *aux* and *et* and schwa monosyllables *je*,[3] *de*, *ne* and *se*.

The final stage of syllable labelling identifies those syllables which are metrically prominent. In work on the meter of English, Halle and Keyser (1971: 169) suggest that stressed syllables which are flanked by unstressed syllables are 'stress maxima'.[4] It is only the position of these stress maxima which is constrained in a syllabo-tonic meter. Adopting this notion, I define prominent syllables as being more stressed than their immediate neighbours:

3. Only <je> is tagged as `s`; spellings suggesting a stressed form (e.g. <gie>, <jou>, <jo>) are tagged as `v` and are thus type 2 monosyllables.

4. 'When a fully stressed syllable occurs between two unstressed syllables in the same syntactic constituent within a line of verse, this syllable is called a "stress maximum"' (Halle and Keyser, 1971: 169). Note that since the notion of prominent syllables is based solely on word stress, I drop the syntactic aspect of this definition.

Prominent syllable rule: Any syllable of a higher value type than its immediate neighbours is prominent (P).

For Halle and Keyser (1971), the stress maximum constraint defines those syllables whose position in the line is determined by a metrical correspondance rule, and Noyer (2002) adopts a similar constraint to test for a syllabo-tonic meter in French verse (cf. chapter one, §4.2). However, even if the verse form does not show a consistent syllabo-tonic meter, I assume that stress maxima are the *most prominent* stressed syllables. The definition is therefore relevant for the present study of MedFr.

The result of applying the prominent syllable rule to the examples from *BelleDame* is as follows:

(4)

	P		P		P		P	
0	2	1	2	1	2	0	2	(0)
je	vueil	lai-	ssier	aux	au-	tres	fai-	re

(*BelleDame*, l. 35)

(5)

	P			P			P
1	2	0	0	2	0	0	2
et	gai-	res	de	bruit	ne	me-	noit

(*BelleDame*, l. 84)

(6)

P							P	
2	0	2	2	2	2	0	2	(0)
car	de	plus	beaulx	jeux	on	se	la-	sse

(*BelleDame*, l. 382)

A number of syllables are left unlabelled by this rule. In (6), there are four unlabelled type 2 syllables, the prominence pattern of which cannot be resolved without resorting to assumptions based on higher-level prosodic constituency. However, the unlabelled syllables in the other examples are clearly less prominent than at least one of their neighbours. To formalize this distinction, 'non-prominent' syllables are also labelled:

Non-prominent syllable rule: Any syllable of a lower value type than one of its immediate neighbours is non-prominent (n).

Applying both rules to the examples, the following final labelling is obtained:

(7)

n	P	n	P	n	P	n	P	
0	2	1	2	1	2	0	2	(0)
je	vueil	lai-	ssier	aux	au-	tres	fai-	re

(*BelleDame*, l. 35)

(8)	n	P	n	n	P	n	n	P
	1	2	0	0	2	0	0	2
	et	gai-	res	de	bruit	ne	me-	noit

(*BelleDame*, l. 84)

(9)	P	n					n	P	
	2	0	2	2	2	2	0	2	(0)
	car	de	plus	beaulx	jeux	on	se	la-	sse

(*BelleDame*, l. 382)

All syllables are now labelled except for the sequence of four type 2 monosyllables in (9). While it would be simpler to analyse the data if all syllables were clearly prominent or non-prominent, we cannot both retain the assumption of regular word stress and propose a prominence labelling for these syllables. In my view, it is preferable to offer a partial labelling based on clear principles than a full labelling which resorts to more debateable or inconsistent assumptions to resolve problematic cases. The methods used to measure the data must therefore be able to treat lines such as (9) which are only partially labelled.

One further consequence of the syllable labelling rules adopted is that type 1 syllables may also be prominent if flanked by type 0 syllables:

(10)			n	P	n	**P**	n	P	
	1	1	1	2	0	**1**	0	2	0
	par	do-	lo-	reu-	se	**sou-**	ve-	nan-	ce

(*BelleDame*, l. 124)

(11)	n	P	n	P	n	**P**	n	P	
	0	2	1	2	0	**1**	0	2	0
	se	moy	ou	aul-	tre	**vous**	re-	gar-	de

(*BelleDame*, l. 237)

Two type 1 syllables are labelled as strong: the initial syllable of the polysyllable *souvenance* in example (10), and the clitic object pronoun *vous* in (11). Any stress realized on these prominent syllables would clearly have been secondary. However, the prominent vs. non-prominent syllable model does not attempt to represent different 'levels' of stress, nor does it make the claim that all prominent syllables bore a primary stress. What the model does claim is that the initial syllable of *souvenance* and the object clitic *vous* are more prominent than their neighbours, thus if a secondary stress were to be realized in either sequence of three syllables, it would be realized on the full vowel rather than on either of the schwa vowels.

1.2 Labelling phonological phrase edges

The reconstruction of PhPh edges in the present chapter is based on the assumption that they may be defined on the basis of syntactic structure. This is the position of Selkirk (1986) and Nespor and Vogel (1986). However, it was shown in chapter one (§1) that rhythmic factors play a role in the delimitation of the ModFr PhPh (cf. Delais-Roussarie, 1995; Post, 2000). For the present study, however, just as syntactic factors were excluded from influencing the labelling of prominent syllables, rhythmic factors such as word length have been excluded from the labelling of prosodic constituent boundaries. This ensures methodological independence in the labelling of prominent syllables and that of PhPh edges.

It also emerged from chapter one that phonological phrasing in ModFr is subject to considerable variation. As such, it will be no more possible to provide a definitive labelling of PhPh boundaries than it was to provide a definitive labelling of prominent and non-prominent syllables. Some word boundaries must remain unlabelled.

The starting point for the analysis is Nespor and Vogel's (1986: 168) definition of the PhPh, reproduced here in Post's (2000) formulation from chapter one:

> **Phonological Phrase formation rule**
>
> A Phonological Phrase groups together a lexical head (X) with all the items on its non-recursive side (i.e. the left) within the maximal projection and with any other non-lexical item on the same side.
>
> (Post, 2000: 34)

In the noun phrase, it is relatively straightforward to apply this rule to MedFr. A PhPh boundary is projected at the right edge of every head noun in the noun phrase. Any function words preceding the head noun are included within the PhPh:

(12) En-la-lance / un-pannon / avoit (*Charrete*, l. 520)

Notationally, I mark labelled non-boundaries with a hyphen, and labelled PhPh edges with an oblique (/). Pre-nominal adjectives, adverbs and non-head nouns are also included in the PhPh:

(13) si t'avrai fet / **molt-grant-bonté,** / (*Charrete*, l. 832)

(14) **mes-sire-Gauvains** / an requoi (*Charrete*, l. 547)

The theory allows for some variation in PhPh formation through PhPh restructuring. Single words which follow the head within the maximal projection, such

as post-nominal adjectives, are optionally included in the PhPh (Nespor and Vogel, 1986: 173). Thus, where a single adjective follows the head noun, the boundary must be left unlabelled:

(15) par-la-resne / a-la-**main senestre, /** (*Charrete*, l. 805)

Notationally, I represent unlabelled boundaries with a single space.

Outside the noun phrase, however, Nespor and Vogel's (1986) definition of the PhPh becomes more difficult to apply. The assumptions made about the syntax of clause structure within the prosodic phonology model are very different to those assumed by the most clearly argued accounts of MedFr clause structure within the generative tradition (Adams, 1987b; Roberts, 1993; Vance, 1997). However, we may still adopt the general principle that only nouns, verbs and adjectives may head a PhPh, while other words cannot. Provisionally, a PhPh edge is marked after all clearly lexical words (nouns, adjectives and non-finite verbs) within the clause. After non-lexical words (conjunctions, the pre-verbal negative, clitic non-subject pronouns) a non-boundary is explicitly marked.

(16) cele **i-**ot fet / **por-son-**repeire / (*Charrete*, l. 738)

(17) bien doit voloir / **qu'-**il fust ocis, /
que-mialz valdroit il morz / **que-**vis: / (*Charrete*, ll. 579–80)

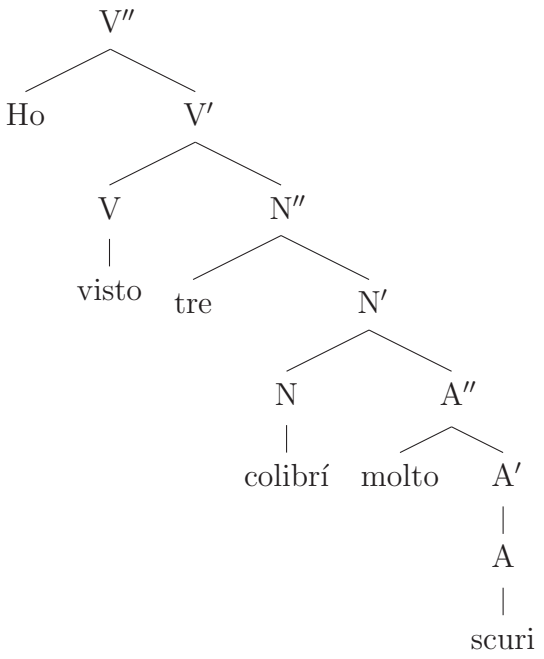
Three types of word boundary are especially problematic to classify: non-clitic pronouns, adverbs and the finite verb. For Nespor and Vogel (1986), only nouns, verbs and adjectives are lexical. The class of words tagged as `Adv` embraces a variety of elements, some of which appear quite functional (e.g. *bien*, *mult*, *plus*, *tant*, *assez*, *si*), but others which are more lexical (e.g. adverbs in *-ment*, which are derived from adjectives). Moreover, there is no clear division in the class. If adverbs in *-ment* are considered lexical since they are derived from adjectives, the same analysis could then be applied to any number of historically multi-morphemic adverbs, which incorporate morphemes derived from lexical categories (e.g. *maintenant*, *tousjours*, *autresi*, *volontiers*). For the present study, I do not attempt to resolve the problem of the phonological phrasing of adverbs, and leave the right edges of adverbial constituents unlabelled. [5]

The phonological phrasing of the finite verb is also problematic. Firstly, it is not clear that Nespor and Vogel's treatment of auxiliaries can be applied to MedFr. The authors assume that where a compound verb form is used, the

5. I return to the issue in part in chapter four (§3.2.3).

past participle is the head of the verb phrase, while the auxiliary verb stands in the same syntactic relation to the participle as the pre-nominal adjective does to the noun:

(18)



'I saw three very dark hummingbirds' (Nespor and Vogel, 1986: 171)

Without anticipating the more detailed discussion of MedFr clause structure in chapter four (§3.1), it is clear that such an assumption is problematic for MedFr. For example, the auxiliary verb may be separated from its participle:

(19) a l'autre fenestre delez
estoit la pucele **venue** (*Charrete*, ll. 544–45)

Moreover, it cannot be the case that the auxiliary verb is simply the specifier of a non-finite head, since the V2 constraint (cf. chapter one, §3.2.1) applies to all finite verbs, whether lexical or auxiliary, but never to non-finite forms. Secondly, there is the potential for the PhPh to be restructured. For example, the finite verb may be immediately followed in the clause by an adverb:

(20) je t'i-voldroie **ja** tenir / (*Charrete*, l. 802)

While it is unlikely that *ja* forms a PhPh in its own right, it is not clear whether it forms the first word in a following PhPh or whether it constitutes the right edge of a restructured PhPh headed by the finite verb, as in ModFr. The most principled choice here is to leave the right edge of the finite verb unlabelled.

The one area where there is sufficient evidence to mark a non-PhPh edge is following pre-verbal subject pronouns, the adverb *si*[6] and the neuter demon-

6. I.e., *si* 'thus', as distinct from the complementizer *si/se/s'* 'if' or reflexive *se/s'*.

strative *ço/ce*. Vance (1995a, 1997) claims that subject pronouns are phonologically proclitic.[7] The most convincing piece of evidence for this analysis is the elision of the schwa-final subject pronoun *je* before a verb or an object pronoun beginning with a vowel:

(21) que **j'en** recevroie tel cop
(*Queste del saint Graal*,[8] p. 197, l. 4; Vance, 1995a: 303)

We assume that elision of this kind may only take place within the same PhPh. A second piece of evidence is the use of combined subject–object pronouns (i.e. the *jel*, *tul* forms, cf. chapter one, §3.2). While we argued in chapter one that such forms were probably lexicalized rather than derived by a productive rule by the 13th century, their existence provides strong evidence that subject and object pronouns must have been within the same PhPh.[9] Moreover, *ço* and *si* share these properties. *Ço* elides before a following vowel:

(22) li chevaliers de la fenestre
conut que **c'estoit** la reïne (*Charrete*, ll. 560–61)

In the 12th century, *si* may serve as a host for consonantal pronouns (as with all such forms, however, examples from later texts are rare):

(23) e **sil** deit cungeer de sei (*LaisMarie*, l. 460)

Moreover, in *Charrete* and *MirNDCoin* in the base corpus, an elided form of *si* is used before a vowel:

(24) **s'avoit** dedanz un chevalier (*Charrete*, l. 553)

Given that both are broadly contemporary Champenois texts, it is possible that the elided *si* form is a dialectal feature. However, this does not imply that the phonological phrasing of *si* was necessarily different in other dialects. From this evidence, we conclude that directly pre-verbal subject pronouns, *si* and *ço* are always incorporated into the following PhPh.

To summarize, the right edges of the following syntactic constituents are labelled as PhPh edges:

– clauses (including parenthetical elements);
– noun phrases;

7. With the important exception of those found in 'conjoined, modified or isolated environments' (Vance, 1995a: 308). Vance considers these forms to belong to a separate series of stressed pronouns, morphologically identical to the unstressed forms.

8. *c.*1220, ed. Pauphilet (1923).

9. The enclitic analysis of the consonantal object pronouns (Horne, 1990a; Dufresne, 1993, 1995; Dufresne and Dupuis, 1994) makes the same assumption, as do the analyses of Skårup (1975) and Vance (1995a).

- head nouns in the noun phrase (unless followed by a single adjective within the noun phrase);
- adjective phrases;
- adjectives (unless pre-nominal in the noun phrase);
- non-finite verbs.

Equally, the following constituents are assumed to be incorporated into the following PhPh, and thus a non-PhPh edge is marked after them:

- conjunctions;
- determiners;
- prepositions;
- pre-verbal non-subject pronouns;
- pre-verbal negative *ne*;
- pre-verbal subject pronouns;
- pre-verbal *si* and *ço/ce*;
- adverbs which are not clausal constituents. [10]

Remaining word edges are left unlabelled. As with prominent syllables, I take the view that it is better to have a principled partial labelling than to attempt to resolve all ambiguous cases.

Below, I reproduce a sample of *Charrete* with the PhPh edges as labelled in the corpus:

(25) Aprés-la-biere / venir / voient
une-rote, / et-devant venoit
uns-granz-chevaliers / qui-menoit
une-bele-dame / a-senestre. / (*Charrete*, ll. 556–59)

1.3 Measures used to present results

1.3.1 The syllable and the pattern method

Labels are assigned to individual syllables. However, when analysing the data, the position of each syllable within the line of verse needs to be taken into account. The simplest technique for doing this is to consider every metrical position individually, and count how many prominent syllables or PhPh edges occur at this position in each extract. This measurement technique I will call the 'syllable' method. Both Le Mée (1978) and Guthrie (1987) used a type of syllable method to generate stress curves, a graphical representation of the

10. E.g. adverbs in the noun or adjective phrase, such as ***molt**-grant-bonté* (13) or adverbs forming part of a complex conjunction, such as *si-que*, *tant-que*.

'average' rhythmic patterns in the text (cf. figures 1, 8, 10 below). These may show, for instance, that even positions in the line are more frequently stressed than odd positions. However, the syllable method ignores interactions within the line. For example, the syllable method may show that even syllables are more frequently stressed than odd syllables, but it is not clear whether this is caused by lines containing only stressed even syllables (i.e. *2–4–6–8*) outnumbering lines containing only stressed odd syllables (e.g. *3–5–8*), or whether a variety of mixed odd and even syllable patterns (e.g. *2–5–8*, *1–4–6–8*) exist in the text.

The pattern method resolves these problems. This method counts the number of lines in a text which are labelled according to a particular pattern, encompassing several metrical positions. For example, we may wish to find the number of lines containing prominent syllables in every even position (a *2–4–6–8* pattern):

(26)

n	P	n	P	n	P	n	P
je	vueil	lai-	ssier	aux	au-	tres	fai-	re

(*BelleDame*, l. 35)

(27)

n	P	n	n	P	n	n	P
et	gai-	res	de	bruit	ne	me-	noit

(*BelleDame*, l. 84)

Example (26) matches the *2–4–6–8* pattern, while (27) does not. However, the pattern method can also measure just a part of the line: for example, only the four central syllables. In this case, example (26) would match a *4–6* pattern; example (27), being stressed on the fifth syllable, would not. In the studies that follow, both syllable and pattern methods are used.

1.3.2 Dealing with unlabelled positions

As discussed in sections 1.1–1.2, both prominent syllable and PhPh edge labelling leave some metrical positions unlabelled.

(28)

		n	P	n	P	n	P	
1	1	1	2	0	1	0	2	0
par	do-	lo-	reu-	se	sou-	ve-	nan-	ce

(*BelleDame*, l. 124)

(29)

		n	P	n		n	P
2	2	1	2	1	1	0	2
qui	ont	es-	poir	d'a-	le-	ge-	ment

(*BelleDame*, l. 26)

Neither line (28) or (29) has a fully defined rhythmic pattern. Rather than exclude valuable data, I have chosen to include lines with unlabelled positions but to weight them according to the certainty with which they match a particular pattern. A fully labelled line (such as 26 or 27) will either perfectly match a pattern or fail to match it. For instance, if all the metrical positions in the line are considered, example (26) matches only a *2–4–6–8* pattern while example (27) matches only a *2–5–8* pattern. Lines with unlabelled positions too can fail to match patterns: for example, both (28) and (29) fail to match the *2–5–8* pattern. However, a line containing unlabelled positions can imperfectly match a variety of patterns. For instance, the labelled positions of both (28) and (29) match a *2–4–6–8* pattern. Both examples would also match a *1–4–6–8* pattern, or a *4–6–8* pattern. Logically, with every unlabelled position, the number of patterns a line could potentially match doubles. Thus, each matching line is assigned a score, where n is the number of unlabelled positions relevant to the pattern:

$$\text{Weighted match score} = \frac{1}{2^n}$$

Example (28) matches the *2–4–6–8* pattern with a score of $\frac{1}{2^2} = \frac{1}{4} = 0.25$, while example (29) matches it with a score of $\frac{1}{2^3} = \frac{1}{8} = 0.125$. A line with no unlabelled syllables has a weighted match score of $\frac{1}{2^0} = \frac{1}{1} = 1$; a line which does not match has a score of 0.

Both pattern and syllable methods use the 'weighted match mean' (WMM) as the unit of measure,[11] giving the mean of all the weighted match scores for a particular pattern in an extract. Because the weighted match score takes account of the ambiguity created by unlabelled positions, the weighted match mean gives the best estimate of the overall proportion of lines in a text which show a particular pattern. The weighted match mean is used throughout the studies unless otherwise stated.

1.3.3 Strength of rhythmic organization

While syllable and pattern methods are useful to track specific changes in rhythmic organization across subcorpora, several measures are produced for each text, particularly in rhythmically diverse octosyllabic verse. When focus-

11. The 'syllable method' is simply a pattern which considers only a single metrical position. Consequently, the same formula is used to calculate the weighted match score of lines in which the syllable in question is unlabelled: i.e. these lines match with a score of $\frac{1}{2^1} = 0.5$.

ing more specifically on differences between texts, it is convenient to represent the strength of rhythmic organization in a text in a single measurement. To calculate the 'strength of rhythmic organization', I take the weighted match means showing the proportion of lines with prominent syllables in positions two, three, four, five and six in the octosyllabic line, and calculate the standard deviation of these five values.[12] A small standard deviation implies that the frequencies of prominent syllables are roughly equal in all of these positions, and is interpreted as showing a lack of rhythmic organization. A high standard deviation implies that prominent syllables in some metrical positions are substantially more common than in others.

1.4 Note on the use of texts

In order to carry out a regression analysis of change over time, and to represent each text graphically on a timeline, the studies in this chapter assign a single year date of composition to each text. Where a dating is expressed as a *terminus a quo* and *terminus ad quem*, the date of composition used in this chapter is the mean of the two.[13] Thus a text dated to the first half of the 12th century is assigned a date of composition of 1125.

In studies 1, 3 and 4, where the effects of dialect and genre are not considered, the early-8 and base subcorpora are combined to give greater chronological coverage of the development of the narrative octosyllable. Regression analyses of change over time are based on the combination of these two subcorpora. In figures, a dotted vertical line at 1170 denotes the chronological divide between the two subcorpora. In study 2, because of the focus on the effects of dialect and genre, the early-8 subcorpus is not used.

12. The final syllable in all lines of verse bears a metrical stress, and as such is ignored in the statistical analyses in this chapter. Regular final prominence also means that the penult cannot be prominent, and this position too is ignored. Note that the weighted match mean for the seventh position is always non-zero, and the weighted match mean for the final position is always less than one. This is due to rare instances in which the two final syllables are unlabelled due to a stress clash. Note also that lines ending in a counted post-tonic schwa syllable in *Brendan* are excluded from studies fd on prominent syllables.

The first position in the line is also ignored. This is not metrically fixed, but has only one neighbouring syllable. A prominence label can be calculated based only on its prominence relative to the second syllable in the line, and this label is used in stress curves. However, since the labelling calculation is different to that of line-internal syllables, I have chosen to exclude the first position from statistical analysis.

13. The date ranges on which this value is based are given in the bibliography.

2 Study 1: Rhythmic change in narrative verse

This study uses data from the early-8, base, ten-syllable and twelve-syllable subcorpora to examine rhythmic change in verse over time. The null hypothesis of the study is that the date of composition of a text is uncorrelated with any rhythmic properties; the hypothesis to be tested is that there is rhythmic organization of word stress in early texts which disappears over time, as suggested by Guthrie (1987) and Noyer (2002). The hypothesis to be tested focuses specifically on the organization of word stress, and as such is based on the labelling of prominent syllables, ignoring the labelling of PhPh edges. Octosyllabic verse from the early-8 and base subcorpora, ten-syllable verse and twelve-syllable verse are examined separately, with general conclusions drawn at the end of the study.

2.1 Octosyllabic verse

2.1.1 Stress curves

The first analysis uses the syllable method to measure the proportion of lines containing a prominent syllable for each metrical position. The weighted match mean is calculated for each of the eight metrical positions in the line for each of the sixteen extracts from the early-8 and base subcorpora.

Stress curves for six of the texts representative of the full corpus are plotted in figure 1. [14] The date given is the single year date of composition assumed in the analyses in this chapter (cf. §1.4). It is instantly apparent that there is a chronological progression. Prominent syllables are substantially more common in the fourth and sixth positions in the line compared to the third and fifth in the earliest texts, a trend which has disappeared by the *BelleDame* extract from the 15$^{\text{th}}$ century.

2.1.2 Change in each metrical position over time

While stress curves give the best impression of the rhythm of an individual text, cross-textual comparison is better shown with a scatter plot. Figure 2 plots the frequency of prominent syllables in each metrical position for each text against the date of composition. For clarity, data for each metrical position is represented in a separate plot, and associated with a different symbol.

14. As is the convention in work on metrics (e.g. Le Mée, 1978; Guthrie, 1987), stress curves are plotted as line graphs.

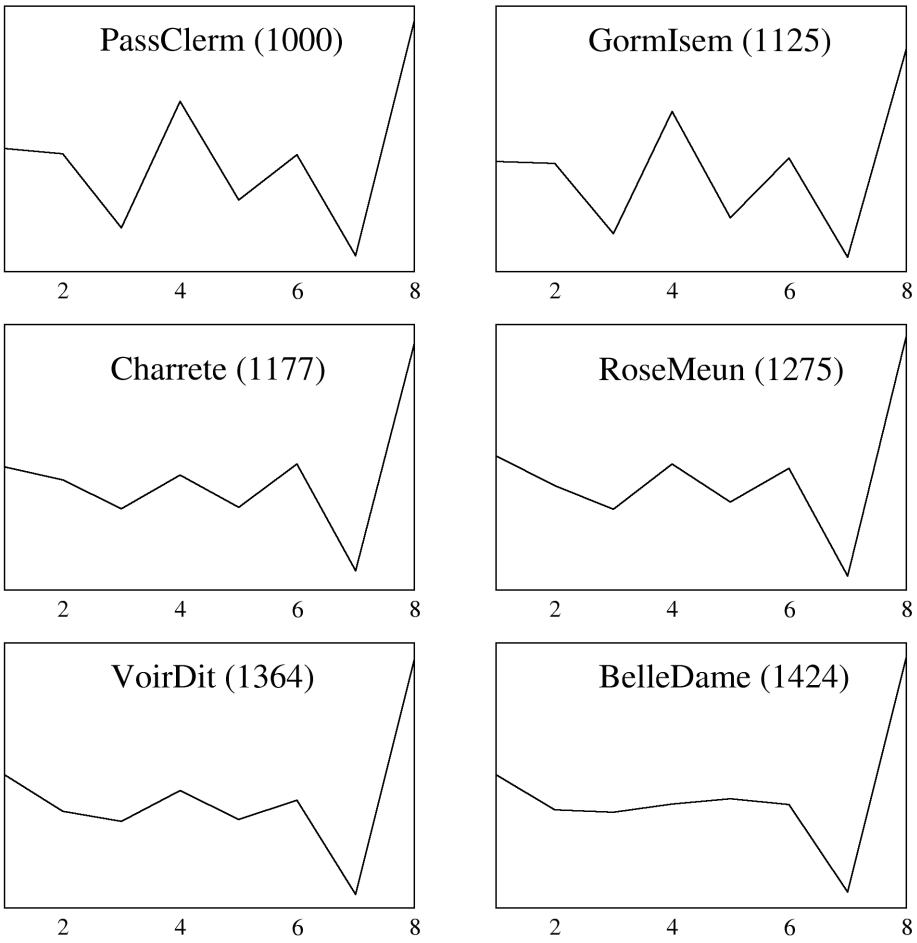


Figure 1: Stress curves for six texts from the early-8 and base subcorpora

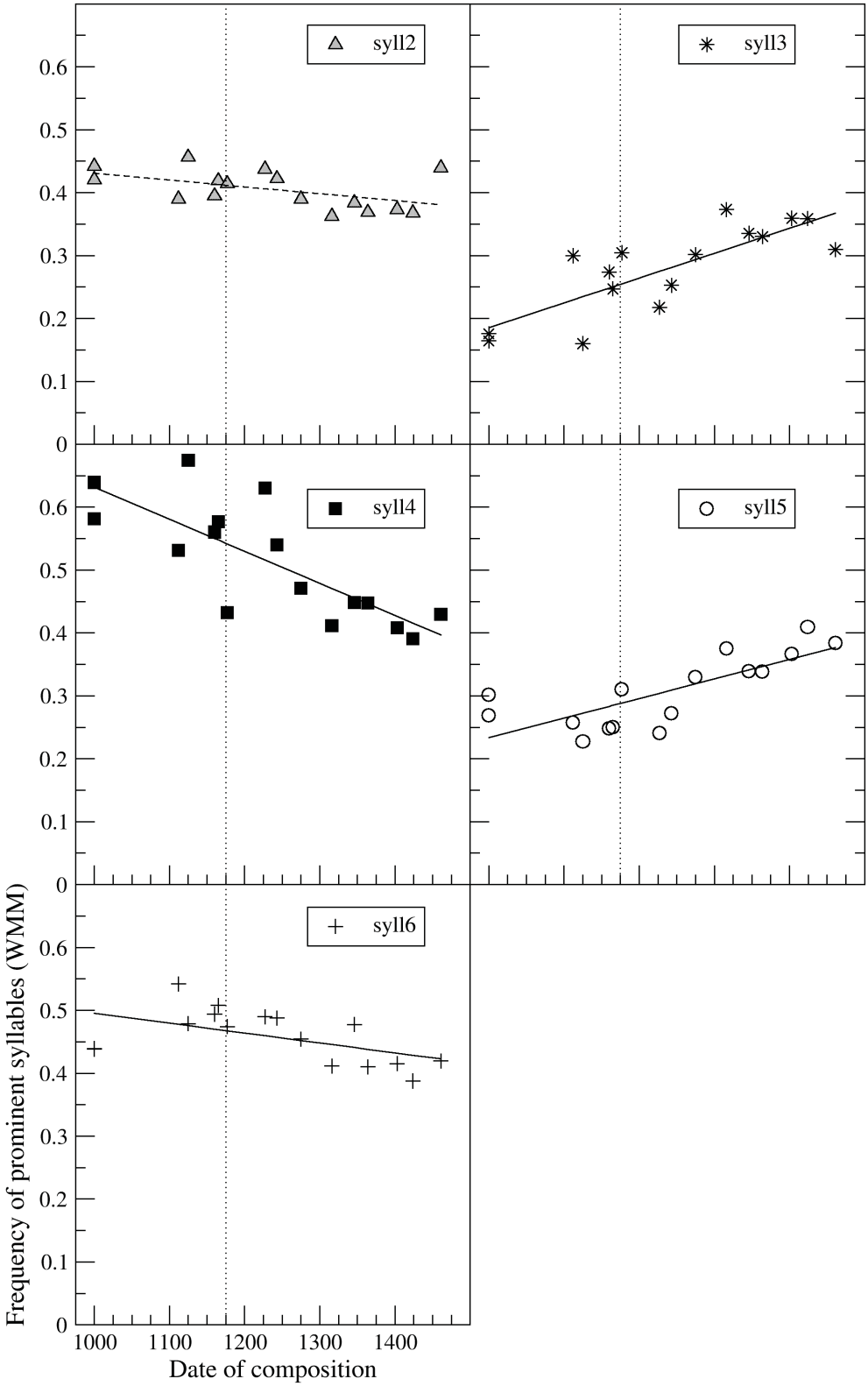


Figure 2: Frequency of prominent syllables in metrical positions two to six, early-8 and base subcorpora

Syllable	r	t	$p > \|t\|$	$\bar{y}$	Slope	Constant
2	0.510	2.217	0.044	0.405	-10.788×10^{-5}	0.539
3	0.804	5.052	0.000	0.279	39.539×10^{-5}	-0.210
4	0.786	4.761	0.000	0.511	-51.009×10^{-5}	1.142
5	0.768	4.481	0.001	0.307	31.053×10^{-5}	-0.077
6	0.525	2.308	0.037	0.458	-15.740×10^{-5}	0.653

Table 1: Regression of frequency of lines with prominent syllables in positions two to six (WMM) on date of composition, early-8 and base corpus

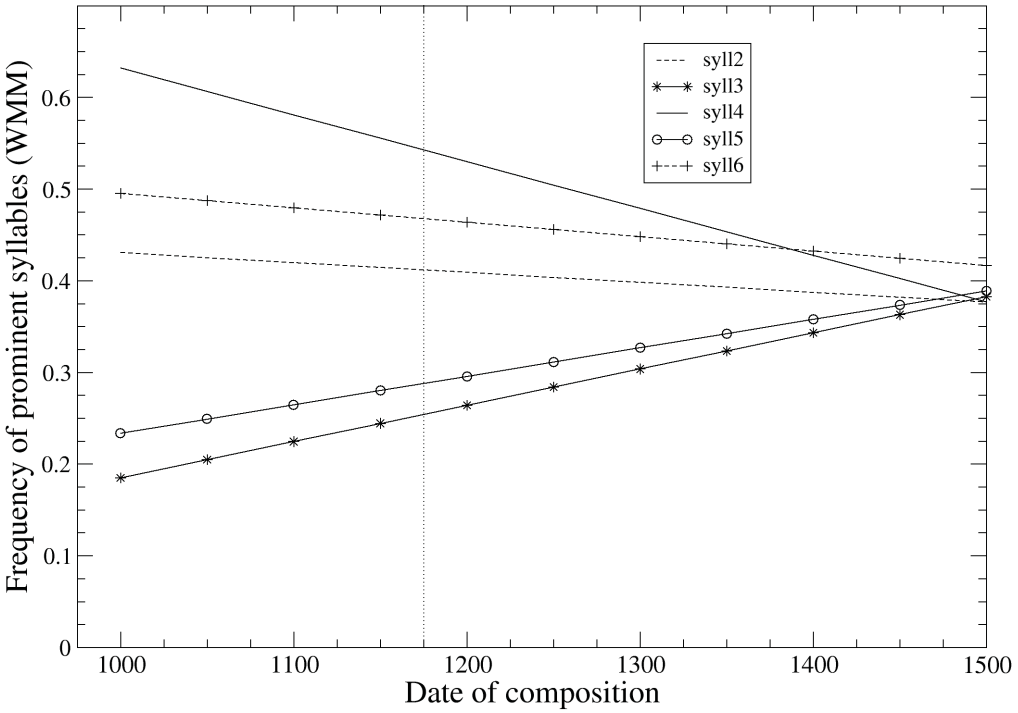


Figure 3: Regression lines for frequency of prominent syllables in metrical positions two to six, early-8 and base subcorpora

A linear regression, using the ordinary least squares method, was calculated for the frequency of prominent syllables in positions two to six of the line on date of composition.[15] This is shown in table 1. Regression lines are plotted together in figure 3. The value of r shows the strength of the correlation between the date of composition of the text and the frequency of prominent

15. Positions one, seven and eight are excluded; see note 12.

syllables in this position: the higher the value of r, the closer to the regression line the actual data points occur. The statistical significance of the correlation, based on a two-tailed t-test, is given in the column $p > |t|$. p-values of less than 0.05 indicate significant diachronic trends in the data at the 95% significance level. Graphically, I use the p-values to draw either unbroken ($p \leq 0.01$) or dashed ($p \leq 0.05$) lines in figures (cf. figures 2 and 3). Additionally, where a trend approaches statistical significance (i.e. $0.10 \leq p < 0.05$), I draw a dotted line in figures: this provides an indication of trends which may be attested if more data were available. Aside from the statistical significance of a trend, the mean frequency of prominent syllables in a particular position across all texts ($\bar{y}$) and the slope coefficient of the regression line give an idea of the linguistic importance of a trend. The slope coefficient indicates the predicted year-on-year rate of change in the frequency of prominent syllables in the given metrical position. Thus, a year-on-year change of -51.009×10^{-5} predicts a -0.26 change in the proportion of lines with a prominent syllable in the fourth position over 500 years, or approximately a 40% decrease from the time of *PassClerm* to that of *Testament*. This is distinct from the strength of the correlation: for example, the correlation between a strong third syllable and date of composition is stronger than that for the fourth syllable, but the slope of the fourth syllable regression line is greater. Strong trends represent a major change in the rhythm of texts; weak trends, while they may be statistically significant, represent only a very subtle change in the rhythm of the line.

From figures 2 and 3 and table 1, we can see that there is a significant trend showing a marked fall in the frequency of prominent fourth syllables, and a corresponding increase in the frequency of prominent third and fifth syllables. Less significant and less marked falls in prominent second and sixth syllables are also attested. Overall, this suggests that the earliest texts have a strong iambic rhythm, which has all but disappeared by the 15$^{\text{th}}$ century. Figure 3 also shows a clear convergence of the frequencies of prominent syllables occurring in each position. By the 14$^{\text{th}}$ century, there is comparatively little difference between the frequency of prominent syllables in the five different positions. More than simply a change in rhythmic organization, this convergence indicates a loss of any rhythmic organization by the 15$^{\text{th}}$ century.

2.1.3 Pattern method analysis

The roughly equal frequency of prominent syllables in individual positions in the 15$^{\text{th}}$ century may conceal some line-by-line regularities. For example,

Pattern	Position				
	2	3	4	5	6
4only	n	n	P	n	n
5only	n	n	n	P	n
2–4	P	n	P	n	n
2–5	P	n	n	P	n
2–6	P	n	n	n	P
3–5	n	P	n	P	n
3–6	n	P	n	n	P
4–6	n	n	P	n	P
2–4–6	P	n	P	n	P

Table 2: Patterns used to measure early-8 and base subcorpora (study 1)

were the *2–4–6–8* pattern, favoured by early texts, to be rivalled in the 15th century by a *3–5–8* pattern, this would appear as a loss of rhythmic organization when analysed by the syllable method.

In using the pattern method to measure the data from the base corpus, I look at nine of the twelve possible rhythmic patterns taking into account positions two to six in the line.[16] The patterns of prominent and non-prominent syllables considered are shown in table 2. Note in particular that the 'only' patterns are not equivalent to a syllable method measurement: the *4only* pattern, for example, requires that positions two, three, five and six do not contain prominent syllables. The results of the analysis are presented in figure 4, with regression analysis in table 3. Regression lines are plotted together in figure 5.

16. Three logically possible combinations are excluded (*2only*, *3only* and *6only*) since they suppose a sequence of four non-prominent syllables. In preliminary versions of this analysis, these were found to be infrequent, and rarely matched exactly by any line. On the exclusion of the first, seventh and eighth position, see note 12.

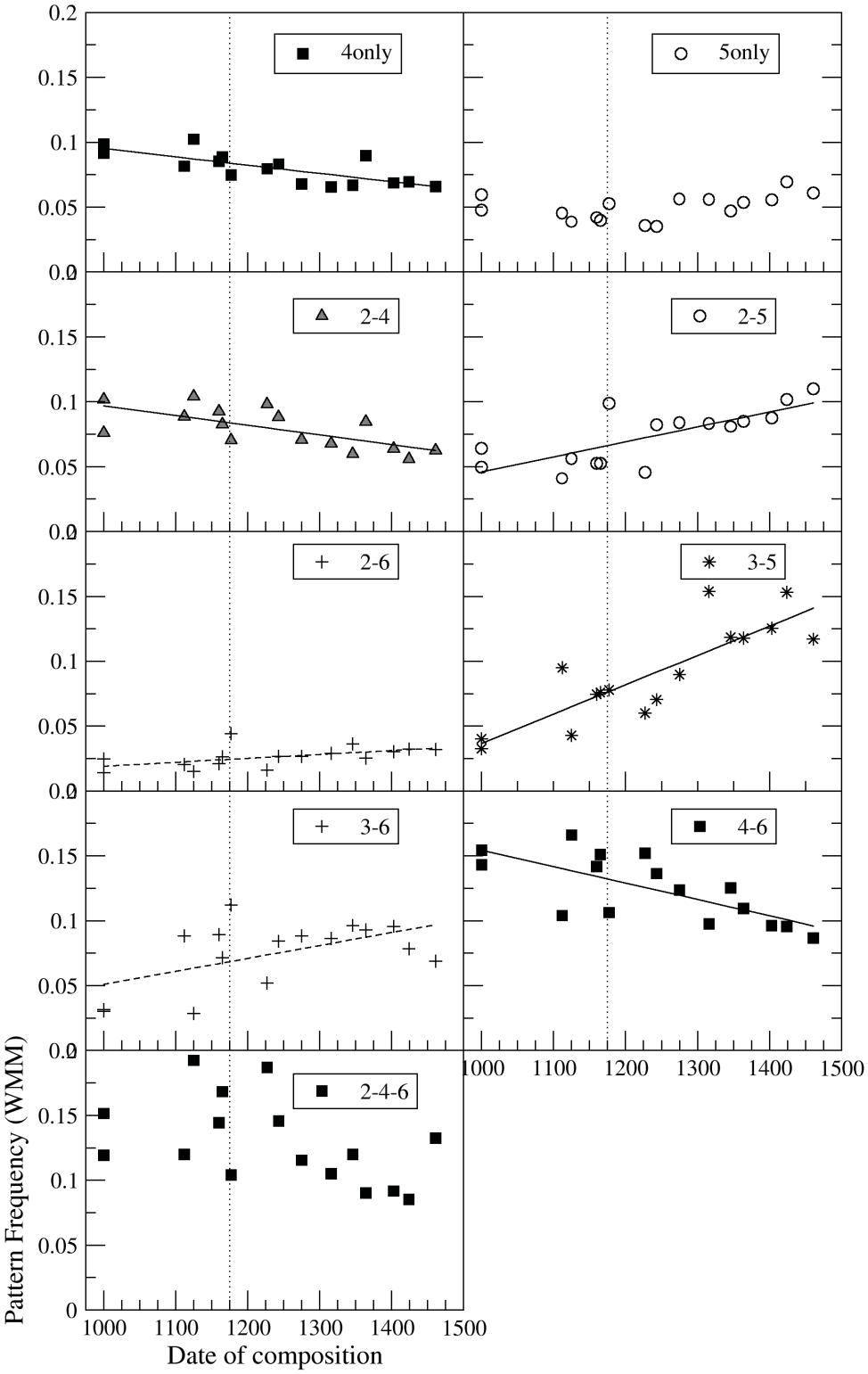


Figure 4: Frequency of individual rhythmic patterns in syllables two to six of the line, early-8 and base subcorpus

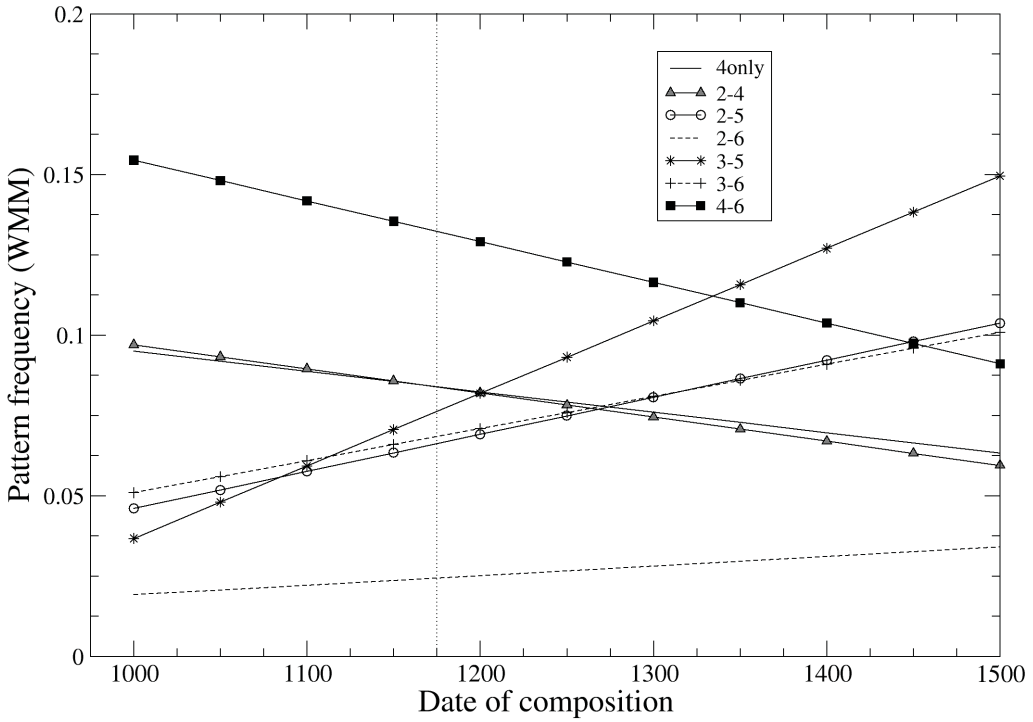


Figure 5: Regression lines for frequency of individual rhythmic patterns in syllables two to six of the line, early-8 and base subcorpus

The data show several statistically significant changes over time. Increases are observed in the *2–5*, *3–5* (99%), *2–6* and *3–6* (95%) patterns over time. Of these, by far the most marked increase is in the *3–5* pattern, with increases in the *2–5* and *3–6* patterns also strong. Significant decreases over time are observed in the *4only*, *2–4* and *4–6* (99%) patterns. Of these, the most marked decrease is in the *4–6* pattern. The overall conclusion reinforces what was observed using the syllable method: patterns containing prominent fourth syllables become less common over time, while patterns containing prominent third and fifth syllables become more common over time. The behaviour of the second and sixth positions is secondary to the three central syllables. For example, the *2–5* pattern becomes more frequent over time, while the *2–4* pattern becomes less frequent; equally, the *3–6* pattern becomes more frequent while the *4–6* pattern becomes less frequent. This is an interesting finding, given the possibility that the iambic rhythmic organization may be related to a tendency to include a mid-line break, as suggested by Guthrie (1987). This is a possibility we will investigate further in study 3.

It does not seem at first glance that the frequency of different rhythmic patterns converges in the 14$^{\text{th}}$ and 15$^{\text{th}}$ centuries, as suggested by the syllable

Pattern	r	t	$p > \|t\|$	$\bar{y}$	Slope	Constant
4only	0.752	4.264	0.001	0.080	-6.37×10^{-5}	0.159
5only	0.436	1.811	0.092	0.050	3.03×10^{-5}	0.012
2–4	0.689	3.554	0.003	0.079	-7.50×10^{-5}	0.172
2–5	0.754	4.293	0.001	0.073	11.54×10^{-5}	−0.069
2–6	0.530	2.337	0.035	0.026	2.97×10^{-5}	−0.011
3–5	0.848	5.998	0.000	0.090	22.57×10^{-5}	−0.189
3–6	0.552	2.477	0.027	0.075	10.00×10^{-5}	−0.049
4–6	0.711	3.785	0.002	0.124	-12.67×10^{-5}	0.281
2–4–6	0.480	2.047	0.060	0.130	-11.15×10^{-5}	0.268

Table 3: Regression of pattern frequency on date of composition, early-8 and base subcorpora

method. However, two of the patterns tested assume a sequence of three non-prominent syllables (*5only*, and *2–6*). Given that the prominence labelling rules may also generate secondary stress, such sequences are rarely attested. [17] Furthermore, since the seventh position is almost invariably non-prominent for metrical reasons, the *4only* and *2–4* patterns also effectively assume a sequence of three non-prominent syllables.

If only the five patterns containing no more than two consecutive non-prominent positions in the line are considered, it is clear that there is convergence in the frequency of different rhythmic patterns by the end of the 13$^{\text{th}}$ century, as shown in figure 6. I have not plotted regression lines in this figure, as these give the impression of gradual change, and in this case change appears to be more abrupt, with a noticeable difference between pre- and post-1250 periods. Few data points fall outside the 0.08–0.13 range from 1250 onwards (only five of 35), whereas the majority of data points from before 1250 fall outside this range (33 of 45). This is clear evidence that no rhythmic pattern is especially favoured in texts from the 14$^{\text{th}}$ and 15$^{\text{th}}$ centuries, while patterns with stress on the fourth syllable are strongly favoured in the 12$^{\text{th}}$ and 13$^{\text{th}}$ centuries.

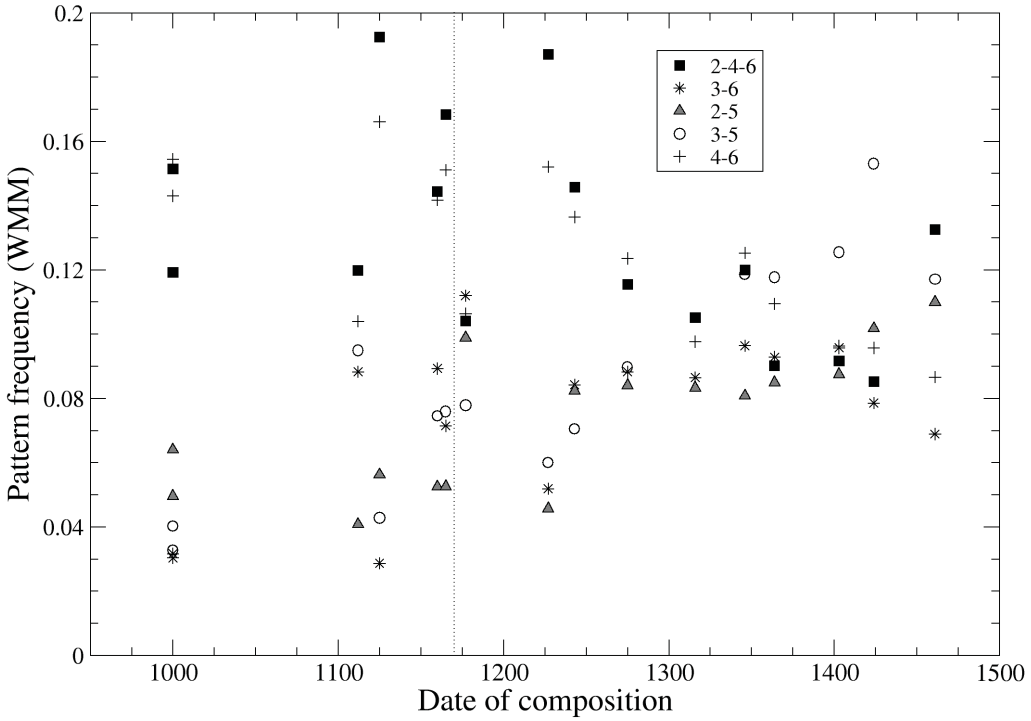


Figure 6: Frequency of all rhythmic patterns containing less than three consecutive non-prominent syllables, syllables two to six of the line, base subcorpus

2.1.4 Variation between texts

So far, we have only considered diachronic trends in the frequency of particular rhythmic patterns across time. Figure 7 focuses more directly on the texts, plotting the strength of rhythmic organization in each text in the early-8 and base subcorpora.

Again, rhythmic organization is shown to be much stronger in earlier than in later texts, with strong rhythmic organization not attested after 1250. However, there is substantial variation between the texts. In particular, *Charrete* shows much weaker rhythmic organization than any other pre-1250 text. Noyer's data corroborates this finding to an extent, noting that in Chrétien's *Yvain* 'rhythmic variety is spread evenly through the line, [and] each line is conceived as a whole undivided by structural cæsura' (2002: 158). Moreover, Noyer suggests that Chrétien's metrical practice 'served as an influential model for future poets' (2002: 159). Yet Noyer's data also includes texts from the

17. Indeed, some analyses of ModFr stress propose to account for the realization of secondary stresses through a rule or constraint restricting the number of unstressed syllables, cf. chapter one, §1.3.

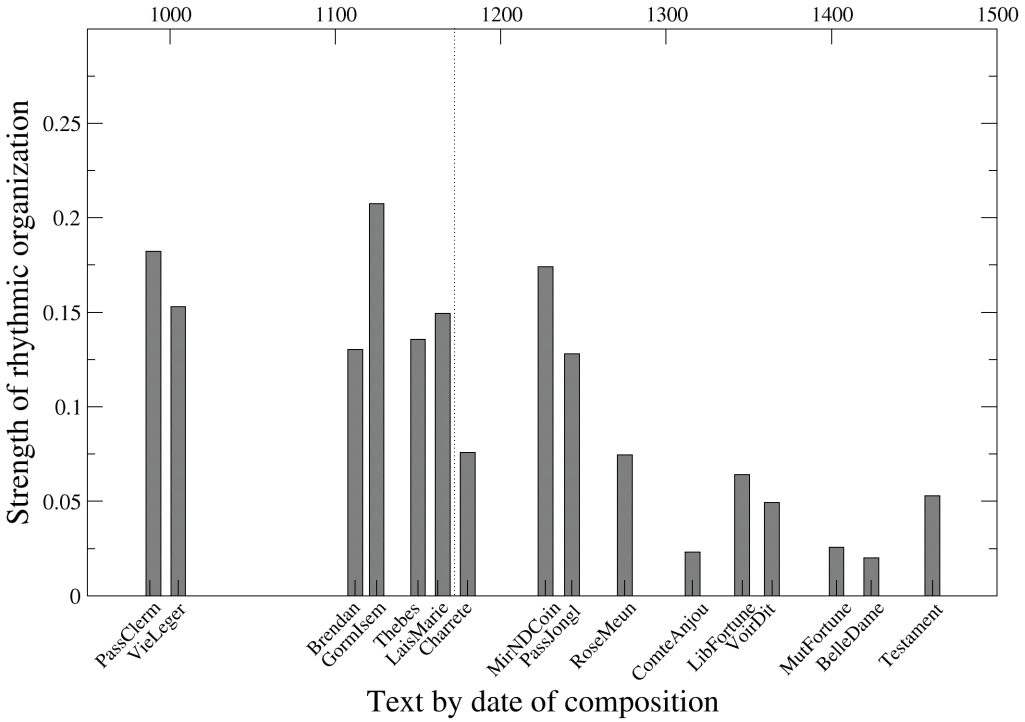


Figure 7: Strength of rhythmic organization, early-8 and base subcorpora

late 12th century which show a similar lack of rhythmic organization: for example, Estienne de Fougères' *Livre des manières* and Wace's *Roman de Rou*. Thus while it possible that Chrétien may have served as a model of rhythmic style for later poets, it is not clear that he was a lone innovator. Note also that the most rhythmically organized text is the only octosyllabic *chanson de geste*, *GormIsem*. I will examine the possibility that some narrative genres show greater rhythmic organization in study 2, and will offer a more detailed analysis of this text in study 3.

Summary

The data suggest that a change in the rhythm of octosyllabic verse took place in the late 12th century and the first half of the 13th century. Texts from the 12th century and earlier generally favour lines with a prominent fourth syllable, often with an iambic rhythm across a whole line. Texts from the 14th and 15th centuries do not show a strong preference for any consistent rhythmic organization in the line, although patterns with prominent fifth syllables are marginally the most common in some 15th century texts. In short, there is a

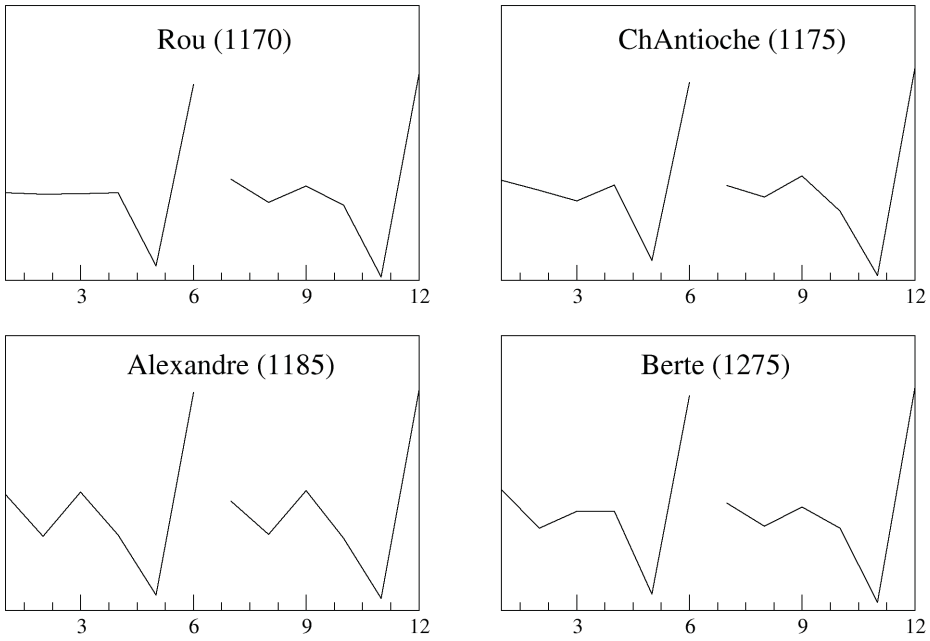


Figure 8: Stress curves for twelve-syllable texts

loss of iambic rhythmic organization in octosyllabic verse, which begins in the 12th century and is complete by the late 13th century.

2.2 Six-syllable hemistichs

2.2.1 Twelve-syllable verse

As with the octosyllabic texts, texts from the twelve-syllable subcorpus were first measured using the syllable method. Stress curves for four of the seven texts are shown in figure 8. All texts use epic cæsura, and therefore metrical restrictions at the cæsura are identical to those at the end of the line (i.e. obligatory stress, and word-final schwa syllables are not counted). Consequently, the first and second hemistichs are labelled as if they were separate lines,[18] and this is represented by a break in the stress curve.

It appears from the stress curves (particularly that of *Rou*) that the two hemistichs show different rhythmic organization. A χ^2 test was performed for

18. I.e. the final syllable of the first hemistich and the first syllable of the second hemistich are not treated as 'neighbouring syllables' by the prominence labelling rules.

Extract	χ^2	p
Rou	7.997	0.046
Alexandre	0.755	0.860
ChAntioche	23.036	0.000
AlexisO	12.923	0.005
Berte	5.114	0.164
HugCapet	8.004	0.046
AlexisA	3.869	0.276

Table 4: χ^2 test on the hypothesis that prominent syllables are differently distributed in first and second hemistich (first four syllables), twelve-syllable corpus (df = 3)

Syllable	r	t	$p > \|t\|$	$\bar{y}$	Slope	Constant
2 (8)	0.609	1.285	0.255	0.374	21.274×10^{-5}	0.109
3 (9)	0.670	2.402	0.061	0.457	-41.738×10^{-5}	0.977
4 (10)	0.598	3.191	0.024	0.349	43.681×10^{-5}	−0.196

Table 5: Regression of frequency of prominent syllables in positions two to four of the second hemistich (positions eight to ten of the line) on date of composition, twelve-syllable subcorpus

each text on the first four syllables of each hemistich to test the hypothesis that the two hemistichs show differing distributions of prominent syllables. The results, presented in table 4, show that this is clearly the case in *ChAntioche* and *AlexisO* (99%), and almost certainly the case in *Rou* and *HugCapet* (95%). In *Alexandre*, *Berte* and *AlexisA* there is no evidence to suggest a difference between the first and second hemistichs. However, the result means that the two hemistichs of the text must be kept distinct in any analysis. A further preliminary test revealed that, except in *Alexis*, no text showed consistent rhythmic organization in the first hemistich. The remainder of this section will focus therefore on the rhythm of the second hemistich.

Figure 9 shows the the frequency of prominent syllables in positions two, three and four of the second hemistich (i.e. positions eight, nine and ten of the line). The regression analysis is given in table 5. Overall, there is a tendency to realize a prominent syllable in the third position in the hemistich, and prominent syllables are most common in this position in every text except for *HugCapet*. Only one significant diachronic trend is attested (at 95%), and that

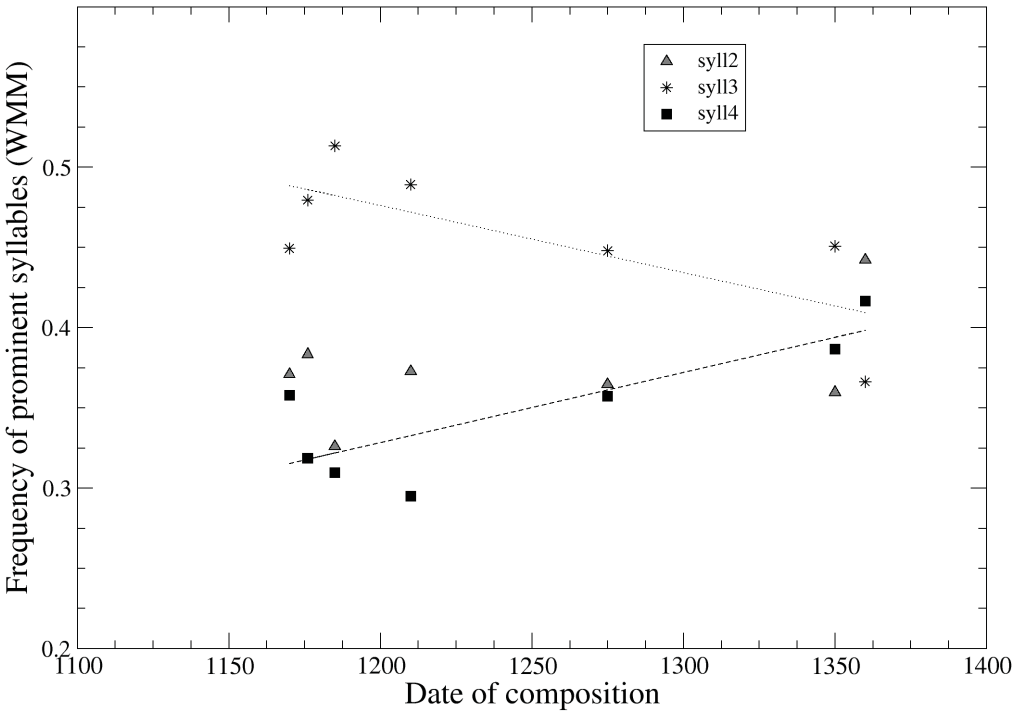


Figure 9: Frequency of prominent syllables, positions two to four of the second hemistich, twelve-syllable subcorpus

is an increase in the proportion of hemistichs containing a prominent fourth syllable.

Figure 9 suggests that as in the octosyllable, later texts tend to show a more even distribution of prominent syllables within the hemistich, while earlier texts show stronger rhythmic organization. Moreover, the most common patterns contain a prominent syllable in the middle of the hemistich, just as in the octosyllable. This is further evidence, perhaps, that rhythmic organization of French verse is a consequence of a more general tendency to include a mid-line (or, in this case, mid-hemistich) prominent syllable. However, diachronic trends are here much less robustly attested. This may be in part because the subcorpus contains fewer texts more unevenly distributed across a shorter time period.

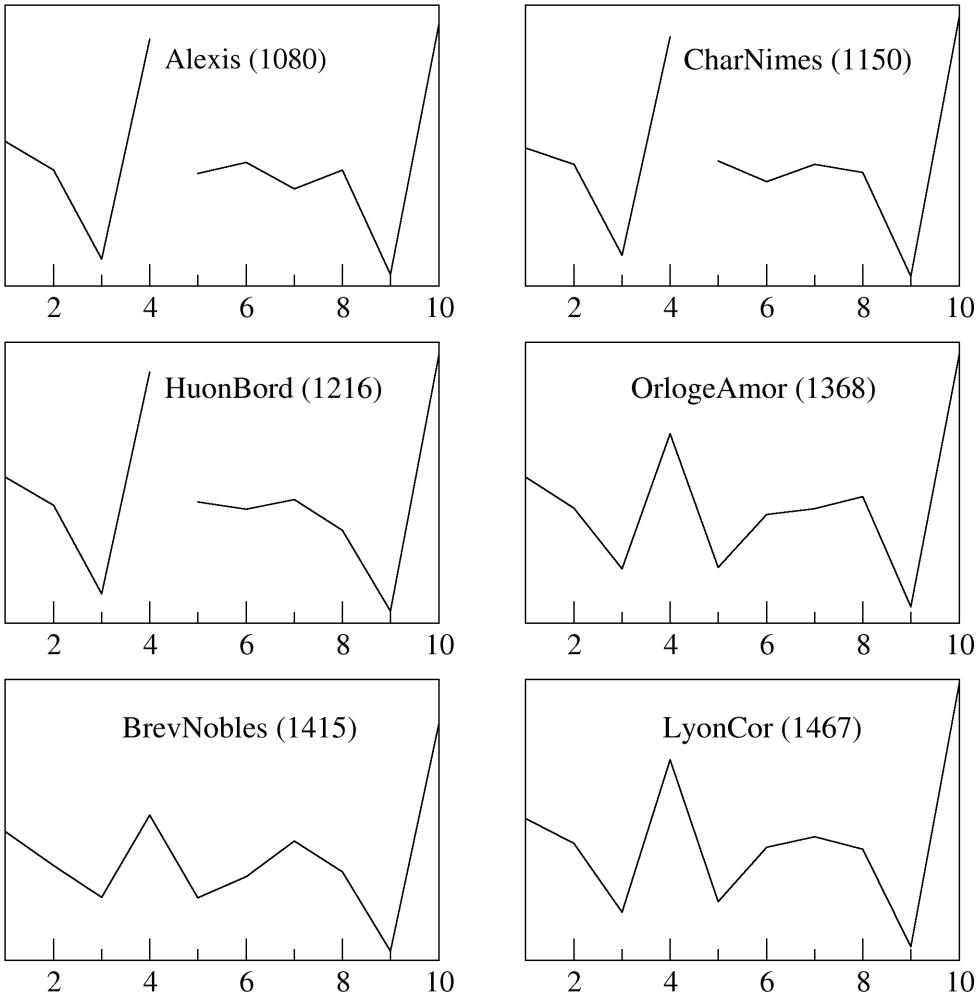


Figure 10: Stress curves for ten-syllable texts

2.2.2 Ten-syllable verse

Ten-syllable verse presents difficulties in analysis not found in either eight- or twelve-syllable verse. Firstly, the epic cæsura of the *chanson de geste*, with a regularly stressed fourth position and any following schwa syllable not counted, ceases to be used in ten-syllable verse from the 14$^{\text{th}}$ century. Later ten-syllable texts adopt the lyric cæsura, in which the third syllable is stressed and the fourth a counted word-final schwa (Elwert, 1965: §89): in the 12$^{\text{th}}$ century, this type of cæsura is only attested in lyric texts. Where texts use epic cæsura, I have labelled the two hemistichs of the line separately as for twelve-syllable verse (cf. *Alexis*, *CharNimes* and *HuonBord* in figure 10). Where texts use

Syllable	r	t	$p > \|t\|$	$\bar{y}$	Slope	Constant
2 (6)	0.524	1.947	0.080	0.392	-31.448×10^{-5}	0.756
3 (7)	0.841	4.911	0.001	0.432	65.429×10^{-5}	−0.325
4 (8)	0.881	5.899	0.000	0.374	-59.920×10^{-5}	1.067

Table 6: Regression of frequency of prominent syllables in positions two to four of the second hemistich (positions six to eight of the line) on date of composition, epic texts from ten-syllable subcorpus

lyric cæsura, the end of the first hemistich no longer has the same metrical properties as the end of the line, and so I have labelled the whole line as one (cf. *OrlogeAmor*, *BrevNobles* and *LyonCor*). [19] Secondly, the cæsura position in ten-syllable verse is occasionally variable, most notably in *OrlogeAmor*, where a 6+4 division may be used (cf. Billy, 1999; and chapter one, note 62). Lines with an unambiguous sixth position cæsura (i.e. with no word boundary after the fourth syllable) are not included for analysis here. However, any line in which a word boundary falls after the fourth syllable is treated as showing a 4+6 division, regardless of whether the word in fourth position is stressed or not.

Stress curves for six of the eleven texts are plotted in figure 10. The fourth syllable is usually prominent, even in those texts where lyric cæsura (with a prominent third syllable) is permitted. The shorter first hemistich has virtually no scope for rhythmic variation, with only the first two syllables not metrically fixed, and the stress curves suggest few changes of interest here. The longer second hemistich does not appear to show strong rhythmic organization in any text, with the possible exception of *BrevNobles*.

Figure 11 shows the frequency of prominent syllables in positions two, three and four of the second hemistich, and is directly comparable to figure 9 above. No statistically significant trends are present in the subcorpus as a whole. However, a different picture emerges if the earliest six texts, with epic cæsura, and the later six texts, with lyric cæsura, are analysed separately. A regression analysis for the epic texts is given in table 6. There are two significant changes (at 99%) in the rhythm of ten-syllable epic texts over the course of the 12$^{\text{th}}$ century. Prominent third syllables become more frequent, while prominent

19. I.e. the last syllable of the first hemistich and the first syllable of the second hemistich are treated as neighbours by the labelling rules.

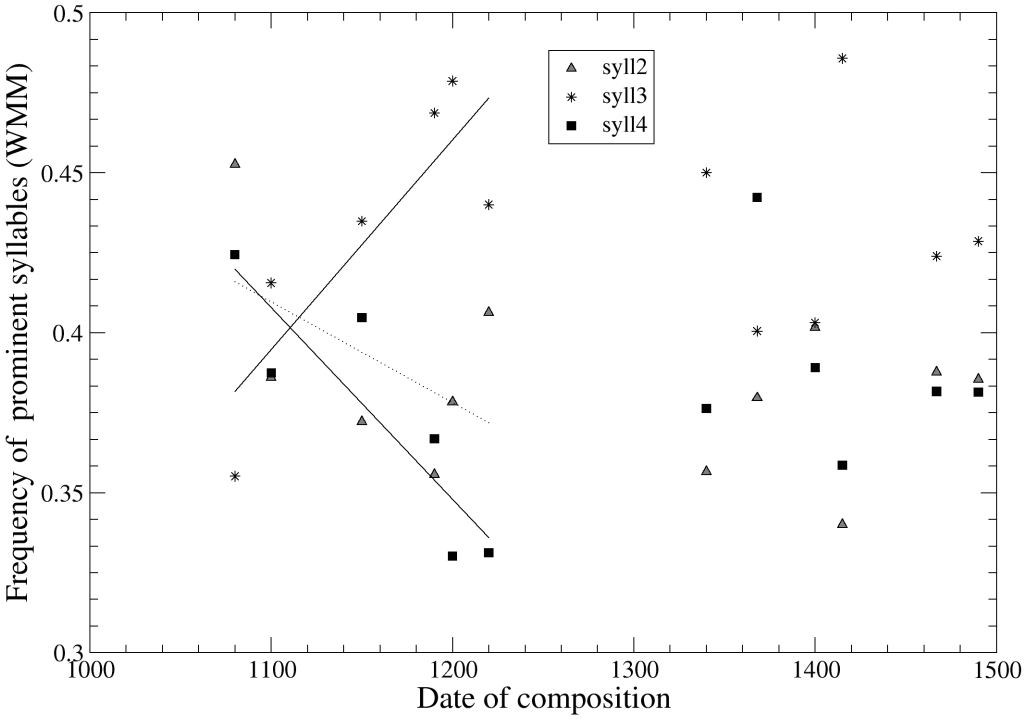


Figure 11: Frequency of prominent syllables in positions two to four in the second hemistich, ten-syllable subcorpus

fourth syllables become less frequent. No significant trends are found in texts with lyric cæsura.

These data show exactly the opposite trend to that observed in the second hemistich of twelve-syllable verse. Moreover, contrary to findings for other verse forms, ten-syllable verse appears to become more rhythmically organized. One possible explanation is that the corpora focus on two different time periods. The epic texts in the ten-syllable subcorpus date from the late 11$^{\text{th}}$ century (*Alexis*) to the early 13$^{\text{th}}$ century (*HuonBord*), while the twelve-syllable texts date from the late 12$^{\text{th}}$ century to the late 14$^{\text{th}}$ century. Thus, the trend in the ten-syllable subcorpus culminates in the early 13$^{\text{th}}$ century with texts showing a preference for a prominent third syllable. By way of contrast, the trend in the twelve-syllable subcorpus *begins* in the last third of the 12$^{\text{th}}$ century with texts showing a preference for a prominent third syllable. It is quite possible that two separate trends affect six-syllable hemistichs in the medieval period: the first, a tendency to adopt a prominent third syllable over the course of the 12$^{\text{th}}$ century, and the second, a tendency to lose strong rhythmic organization

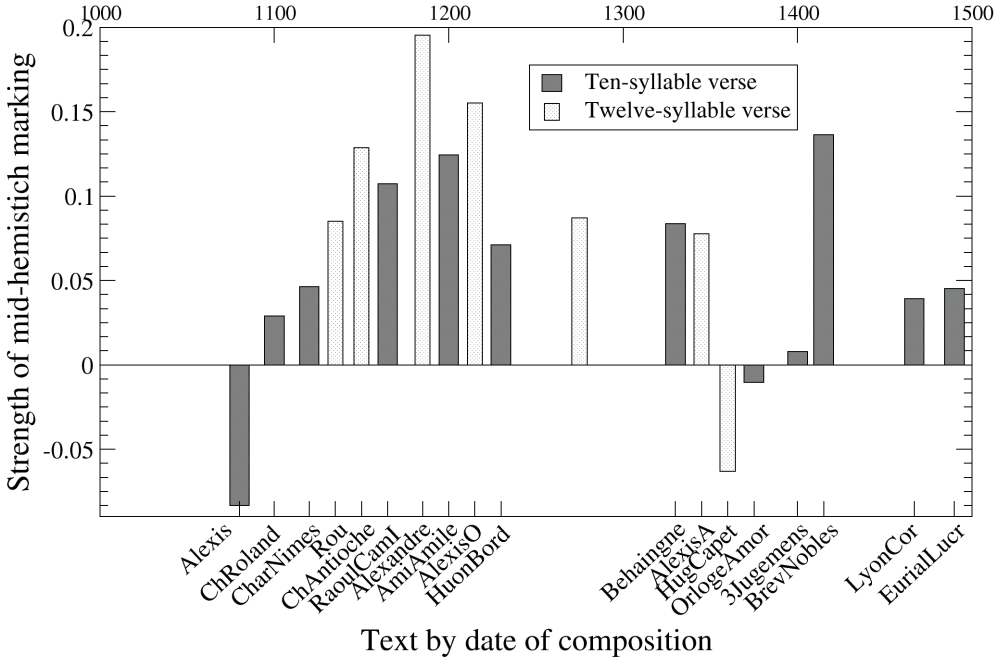


Figure 12: Strength of mid-hemistich marking, second hemistich, ten- and twelve-syllable subcorpora

from the 13th century onwards. This explanation finds some support in the ten-syllable texts: four of the six lyric texts show a lower frequency of the *syll3* pattern than any of the mid to late 12th century epics.

2.2.3 Variation between texts

Figure 12 captures the rhythmic variety of the second hemistich of each ten- and twelve-syllable text in a single statistic. The mean frequency of prominent syllables in the second and fourth positions of the hemistich is subtracted from the frequency of lines with a prominent syllable in third position. Texts with a strong positive value (e.g. *Alexandre*, *RaoulCamI*, *AmiAmile*, *AlexisO*, *BrevNobles*) show a preference for a prominent third syllable; texts with a strong negative value (*Alexis*, *HugCapet*) show a preference for iambic organization. Of the ten-syllable texts, only the earliest (*Alexis*) shows a preference for iambic rhythmic organization. [20] Other metrical traditions using an compa-

20. Klausenburger (1970: 25–27) suggests a 75% correspondance of word stress with even positions in this text, and compares this with 69% in *ChRoland*. However, I have serious reservations about the methodology of this study. Firstly, metrically fixed stresses in fourth

rable verse form favour iambic lines. Most notably, Chaucer and Gower employ a regular iambic rhythm in their English ten-syllable versification, and Gower extends this to his French verse (Duffell, 2008: 87–92). Italian verse of the 14th century also shows a marked preference for iambic lines: Duffell (2008: ch.4, note 28) notes that 79% of lines in Petrarch and 70% in Boccaccio are iambic, while variants with a 4–7–10 stress pattern (the equivalent of a French line with a prominent third syllable in the second hemistich) account for only 5% of Petrarch's lines and 15% of Boccaccio's. Based on this evidence, it is tempting to suggest that iambic organization in particular is favoured in languages with strong word stress.

Yet where French ten- or twelve-syllable texts show a rhythmic preference, it is almost invariably for a mid-hemistich division. Chronologically, this preference is most notable in texts from the late 12th century and the early 13th century. Figure 12 illustrates why no trends could be identified in texts with a lyric cæsura: there is a high degree of variation between texts, with some (*Behaingne*, *BrevNobles*) favouring a 3+3 rhythmic organization, while others (*OrlogeAmor*, *3Jugemens*) show no preference. In the case of *OrlogeAmor*, influence from Italian versification practice in the use of sixth position cæsura may be reflected in the slight preference shown for even syllable prominence in the second half of the line. Two texts, however, show atypical rhythmic patterns. Most notably, *HugCapet* from the mid-14th century is the only text apart from *Alexis* to show a preference for iambic organization. The text is a late *chanson de geste*, and as one of the last sample of the genre it would be no surprise if it were rhythmically conservative. However, even *ChRoland* and *CharNimes* do not favour iambic organization of the six-syllable hemistich to this extent. The strong preference for a prominent third syllable in *BrevNobles* is puzzling, especially since Chartier's octosyllabic *BelleDame* showed no strong rhythmic organization. In study 4 (§5.2), I will discuss a finding from the rhythm of Alain Chartier's prose which suggests that the mid-hemistich stress of *BrevNobles* is not necessarily caused by deliberate rhythmic organization.

2.3 Conclusions from study 1

In this study, clear evidence for rhythmic change in MedFr verse over time has been found. The change may be characterized as a loss of rhythmic or-

and tenth position are included in the count, thus 40% correspondance is guaranteed. Secondly, the brief example cited contains the dubious stress reconstruction: *Bons fut li sec.(les) al tens an.cï.e.nur* (1970: 25–26). This places stress on the copular *fut* rather than the fronted adjective *bons*.

ganization over the course of the 13th century, with a turning point in the octosyllabic subcorpora at around 1250. Evidence for rhythmic organization of word stress supports the hypothesis that stress deafness had not emerged by the early 12th century, but was well-established by the mid-13th century. The change is most clearly observed in octosyllabic texts: a preference for rhythmic patterns with prominent even syllables, particularly the mid-line fourth syllable, is present in the majority of texts from the 11th and 12th centuries, but has disappeared by the 14th and 15th centuries. However, change is less clearly attested in the six-syllable hemistichs of ten- and twelve-syllable verse. A tendency to use iambic rhythmic organization is found in the earliest ten-syllable text, *Alexis*. Over the course of the 12th century, hemistichs with a prominent third syllable become more common, and despite a slight decline in the 13th and 14th centuries, this remains the most common rhythmic organization of six-syllable hemistichs throughout the medieval period. Thus, it appears that loss of rhythmic organization is observed most strongly where iambic rhythm is historically favoured.

However, the data suggest that two further hypotheses must be investigated before we can confidently attributed change in versification to a linguistic change. Firstly, the iambic rhythm of pre-1250 octosyllabic verse is characterized primarily by a prominent fourth syllable. In study 3, I return to the hypothesis that the prominent fourth syllable of octosyllabic verse is associated with the edge of a prosodic constituent, and thus the rhythmic organization of the line is not purely dependent on the position of word stress. Secondly, while overall diachronic trends are attested, significant inter-textual variation has also been observed. This raises the possibility that the apparently gradual diachronic trends mask a more radical change from one style of versification to another, which in turn may be a consequence of the development of different genres in the medieval period. If so, then examining textual parameters to discover among which authors and in which text types this variation first appears will be essential.

3 Study 2: The effects of dialect and text type on rhythm

This study tests two separate hypotheses relating to the rhythmic change in octosyllabic verse observed in study 1. The first hypothesis is that rhythmic change is affected by dialectal factors. In chapter two (§1), it was suggested

that contact with Germanic made eastern, northern and Anglo-Norman dialects more likely to retain word stress, and hence syllabo-tonic rhythmic organization in verse, for longer. The second hypothesis is that the rhythm of verse varies according to text type.

3.1 Dialect

This section of the study uses texts from the three dialect subcorpora alongside texts from the base subcorpus composed between 1175 and 1300 (the 'dialectal' subcorpora). The short time period over which dialectal differences are analysed coincides with the time period at which the critical rhythmic change in the base corpus takes place (cf. study 1), and as such will give a clear idea as to whether the same change is present in the same form in all dialects.

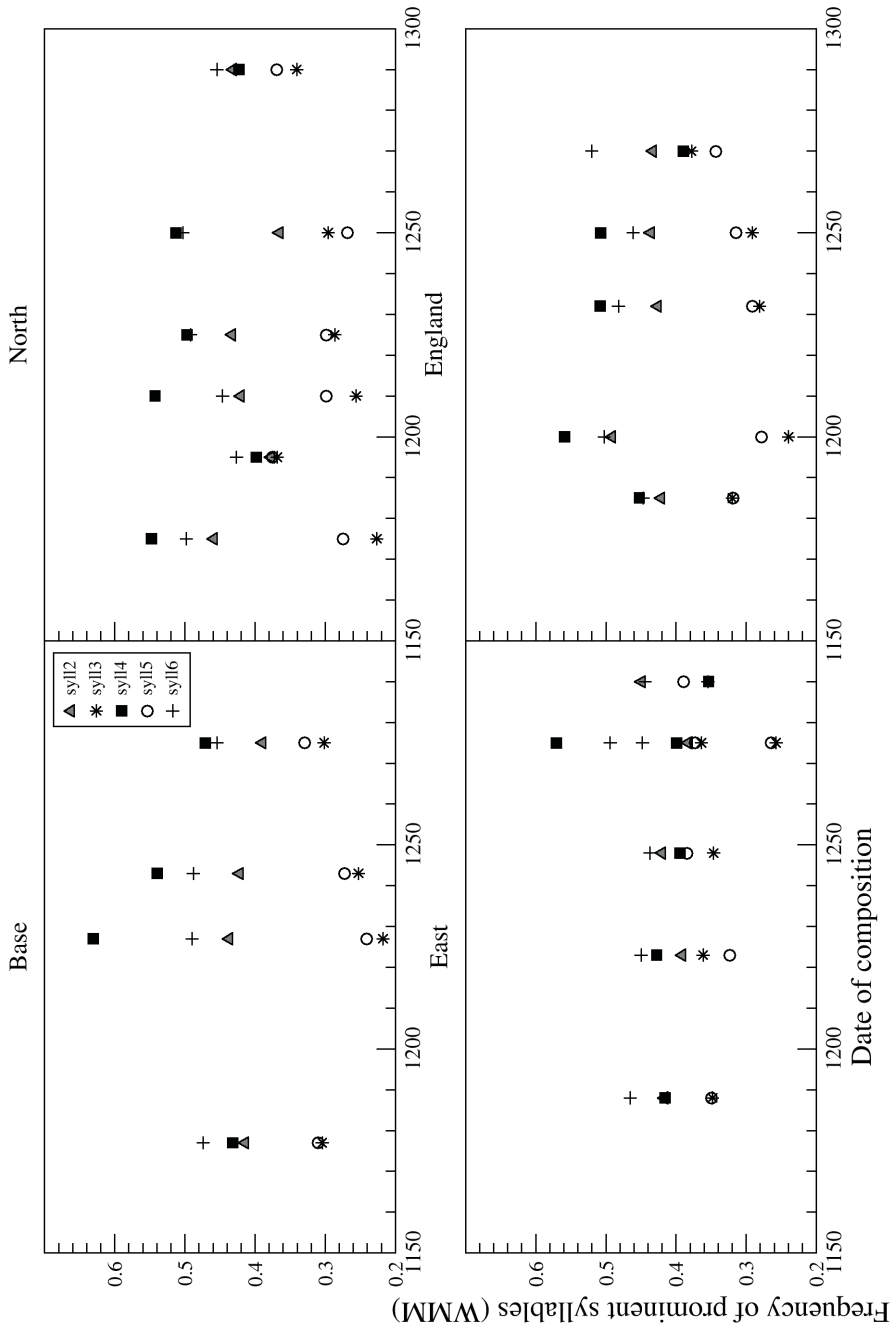


Figure 13: Frequency of prominent syllables in dialectal subcorpora

Syllable	r	t	$p > \|t\|$	$\bar{y}$	Slope
2	0.261	1.177	0.254	0.419	−0.017
3	0.444	2.160	0.044	0.304	0.048
4	0.421	2.023	0.057	0.475	−0.067
5	0.440	2.135	0.046	0.317	0.042
6	0.348	1.616	0.123	0.470	−0.020

Table 7: Regression of frequency of prominent syllables in positions two to six of the line on [±Eastern], dialectal subcorpora

3.1.1 Quantitative overview

The data were first measured using the syllable method. Figure 13 shows the frequency of prominent syllables in positions two to six in the line plotted separately for each of the four dialects studied. Few trends in the data are immediately obvious. Indeed, the only significant dialectal difference attested in the data is shown in table 7. Compared with the other texts in the dialectal corpus, eastern texts tend to have prominent third or fifth syllables (just significant at 95%) and are perhaps less likely to have prominent fourth syllables (almost significant at 95%). The difference is not huge: the frequency of prominent syllables in positions three to five differs by 12–14% in the eastern texts from the general mean.[21] This implies that eastern texts are less rhythmically organized than those written in other dialects. Such a finding is unexpected, since we had suggested that contact with Germanic was likely to make eastern, northern and possibly Anglo-Norman texts more likely to show word stress and rhythmic organization than their central counterparts. Instead, there is no significant difference between central texts and the others, and it is eastern texts which apparently show less rhythmic organization.

Regression analyses on other dialectal variables failed to show any further significant correlation. Nor is there any correlation of rhythmic organization with date of composition. This is true both when all the texts are considered together, irrespective of dialect, and when the individual subcorpora are analysed separately. We conclude that over this comparatively short period of time (125 years), it is not possible to distinguish the same long-term changes

21. Given by dividing the slope coefficient by the mean ($\bar{y}$). As [±Eastern] is a binary (or 'dummy') variable, the slope value here simply indicates the projected difference between eastern and non-eastern texts.

in the rhythm of verse that are apparent over a longer time period. This is probably due to a combination of factors, including substantial variation between individual texts, the relatively slow rate of long-term change observed in study 1, and the fact that inaccuracies in the date of composition become proportionally larger when a shorter time period is investigated.

Two further points need to be made. Firstly, although there is no evidence that octosyllabic lines in Anglo-Norman texts show greater rhythmic organization than their continental counterparts, the syllable count of 13th-century Anglo-Norman texts is very irregular (cf. chapter two, §2.2.3). We may conclude that while the versification of 13th-century Anglo-Norman texts is clearly different, there is no evidence from the present study that it is based on regular rhythmic organization. Secondly, since the finding from eastern texts is entirely contrary to the hypothesis, I suspect that other factors, such as the genre of the texts chosen, may have affected the result. This possibility is considered in the following section.

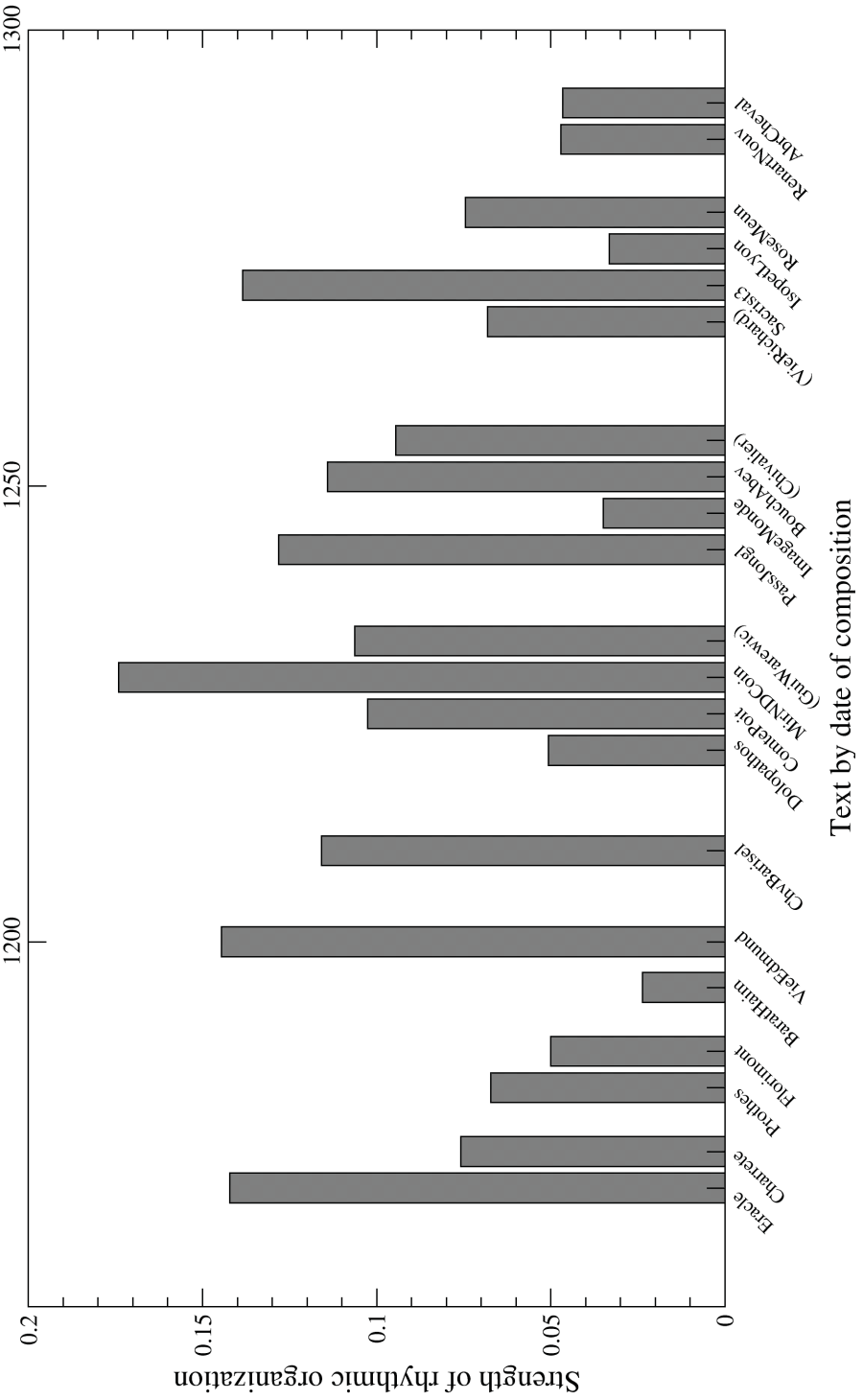


Figure 14: Strength of rhythmic organization, dialectal subcorpora

3.1.2 Variation between texts

Figure 14 shows the strength of rhythmic organization in each of the extracts in the dialectal subcorpora. Names of Anglo-Norman texts with irregular syllable counts are given in parentheses. It is clear that texts with little rhythmic organization (e.g. *BaratHaim*, *ImageMonde* and *IsopetLyon*) are attested throughout the time period, as are texts with strong rhythmic organization (e.g. *Eracle*, *MirNDCoin* and *Sacrist3*). All texts are narrative. However, one pattern which emerges is that the majority of romances (*Charrete*, *Prothes*, *Florimont*, *Dolopathos*, *RoseMeun*, *RenartNouv*) show less rhythmic organization than other texts of the same period, with the exception of the early *Eracle*, *ComtePoit* and the irregular Anglo-Norman *GuiWarewic*. Among the most strongly rhythmically organized texts, we note two of the four *fabliaux* (*Sacrist3* and *BouchAbev*), although another, Jean Bodel's *BaratHaim*, shows the least rhythmic organization of any pre-1250 text. However, Guthrie (1987) observes a similar lack of rhythmic organization in Jean's lyric verse, commenting that Jean is an 'obvious maverick' (1987: 64). The versification of *BaratHaim* may thus be atypical of the *fabliau* genre. The text with the strongest rhythmic organization overall is *MirNDCoin*. While the text has a religious subject matter, it is in some respects similar to the *fabliaux*. Beaussart (1989), for example, notes parallels in the narrative structure:

> Le schéma narratif se révèle lui aussi d'une grande simplicité et il est bien difficile de ne pas le rapprocher de celui du conte merveilleux.
>
> (Beaussart, 1989: 18)

While no exceptionless generalizations can be made, the picture emerging from figure 14 is one in which more popular genres strongly associated with oral performance show relatively strong rhythmic organization, while in romances, clerical compositions in which the spoken dimension was perhaps less important, rhythmic organization is less common. In the eastern subcorpus, the only text to show strong rhythmic organization is a *fabliau* (*Sacrist3*); the remaining texts are less clearly associated with spoken performance (two romances, one didactic piece and a translation of Æsop's *Fables*). This raises the possibility that the rhythm of medieval texts may be influenced by their context of communication, with those having the strongest links to oral performance showing the strongest tendency to rhythmic organization. I will return to this possibility in section 3.3.

Pattern	r	t	$p > \|t\|$	$\bar{y}$	Slope	Constant
NARRATIVE						
4only	0.586	2.286	0.045	0.075	-5.221×10^{-5}	0.048
5only	0.728	3.354	0.007	0.050	7.323×10^{-5}	0.086
2–4	0.686	2.979	0.014	0.075	-9.134×10^{-5}	0.028
2–5	0.695	3.053	0.012	0.080	13.467×10^{-5}	0.142
3–5	0.787	4.038	0.002	0.103	24.335×10^{-5}	0.307
4–6	0.779	3.925	0.003	0.119	-17.112×10^{-5}	0.027
2–4–6	0.578	2.238	0.049	0.124	-17.789×10^{-5}	0.016
LYRIC						
2–5	0.506	1.856	0.093	0.088	9.227×10^{-5}	0.096
4–6	0.568	2.185	0.054	0.101	-15.306×10^{-5}	0.077
THEATRE						
4only	0.560	2.026	0.073	0.063	-4.740×10^{-5}	0.064
3–5	0.576	2.113	0.064	0.116	16.470×10^{-5}	0.232
4–6	0.539	1.920	0.087	0.099	-12.648×10^{-5}	0.031
2–4–6	0.537	1.909	0.089	0.101	-8.889×10^{-5}	0.068

Table 8: Regression of pattern frequency on date of composition, text type subcorpora, all significant results and those approaching significance

3.2 Text type

This section of the study uses data from the base subcorpus, the theatre subcorpus, and octosyllabic sections of the lyric subcorpus (the 'text type' subcorpora). Since dialectal differences were shown in the previous section to be of little significance, the two texts dating from the second half of the 12$^{\text{th}}$ century from the early-8 subcorpus (*Thebes* and *LaisMarie*) are also included. These texts are contemporary to the earliest theatrical text (*JeuAdam*), and ensure that more than a single narrative text (*Charrete*) is included for this period.

3.2.1 Quantitative overview

Following a preliminary investigation using the syllable method, the pattern method was used to measure the data, using the same patterns as those used on the base corpus in study 1 (see table 2). Figure 15 plots the frequency of the five patterns with sequences of no more than two non-prominent syllables. A regression analysis of pattern frequency on date of composition for each of the three subcorpora is presented in table 8.

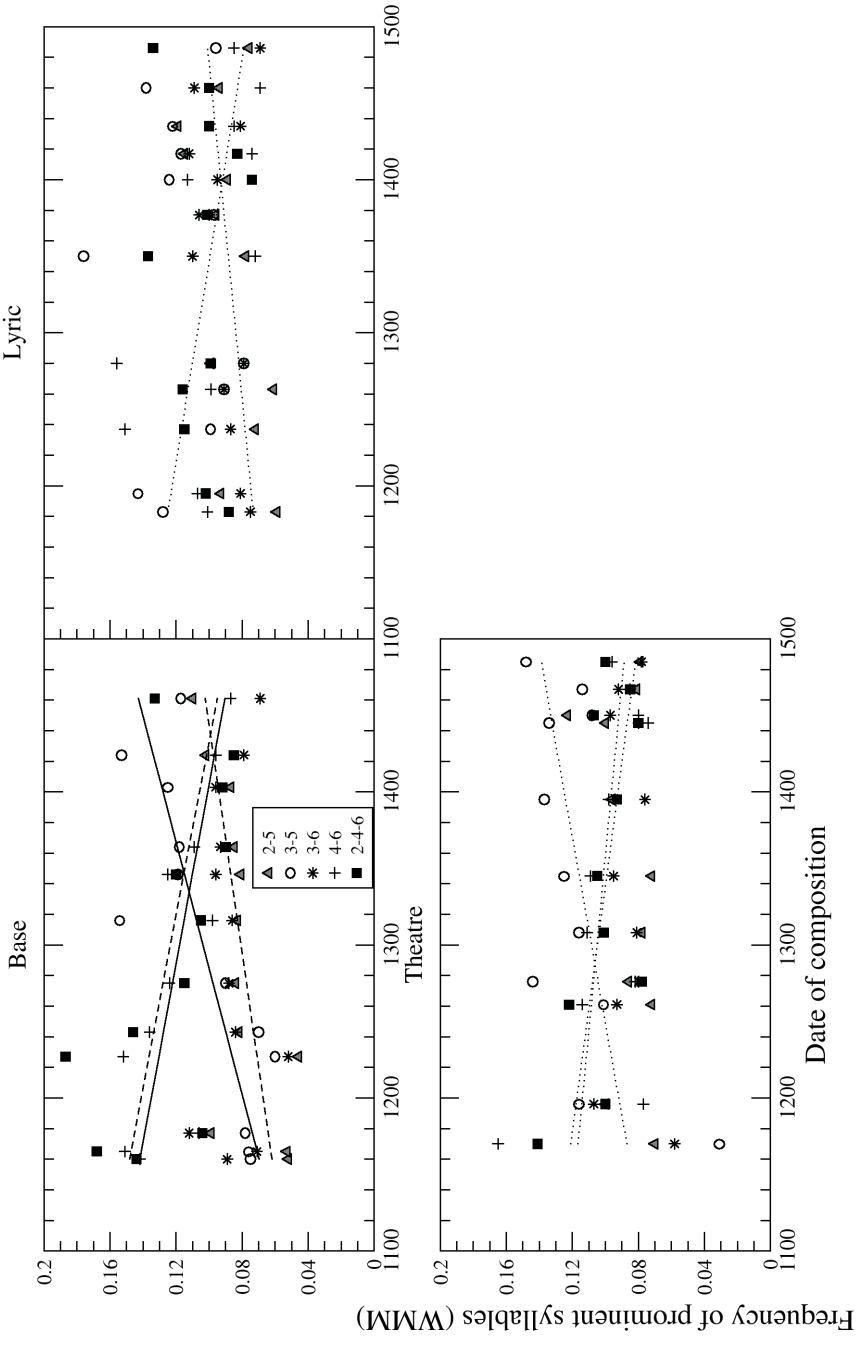


Figure 15: Frequency of all rhythmic patterns containing less than three consecutive non-prominent syllables, syllables two to six of the line, text type subcorpora

Significant change over time is only attested in the base subcorpus. As discussed in study 1, here there is a significant (at 95%) decline in patterns with a prominent fourth syllable (*4only*, *2–4*, *4–6*, *2–4–6*), while there is a corresponding rise in patterns with prominent third and fifth syllables (*5only*, *2–5*, *3–5*). The plot for the base corpus in figure 15 thus shows the convergence of the frequency of different rhythmic patterns in the 14th and 15th centuries first observed in study 1, with perhaps patterns containing prominent fifth syllables becoming slightly more common in many 15th-century texts. In the other two subcorpora, no significant change is attested. However, a number of trends approach significance, and these all parallel significant trends attested in the base subcorpus. Lyric texts show a decline in the prominent fourth syllable *4–6* pattern which approaches significance. The weakness of this trend is perhaps explained by the fact that the two lyric texts from the late 12th century do not favour patterns with a prominent fourth syllable. Indeed, the most common rhythmic pattern in the two earliest lyric extracts (*Blondel* and *GaceBrule*) is the *3–5* pattern. Narrative texts in which this pattern is the most frequent are not attested until the 14th century. In theatrical texts, there is an increase in the proportion of patterns with prominent third and fifth syllables at the expense of prominent fourth syllables. Trends approaching significance are found in the *4only*, *4–6* and *2–4–6* patterns, which all become less frequent, while the *3only* and *3–5* patterns become more frequent. However, it should be noted that the earliest text (*JeuAdam*) shows a strong preference for prominent fourth syllables, and thus it is difficult to establish whether there is true change over time in this text type, or whether in fact an atypical early text (*JeuAdam*) is having a disproportinate effect on the analysis. Particularly remarkable in the theatre subcorpus is the early emergence of the *3–5* pattern as the most common. In eight of the ten texts from 1200 onwards, this pattern is preferred. In study 1, I suggested that a convergence of rhythmic patterns was illustrated by the fact that the majority of data points from patterns with sequences of no more than two non-prominent syllables in the post-1250 period cam within the range 0.08–0.13. Taken as a whole, only 26 of 115 data points from the whole of the theatre and lyric subcorpora exceed these limits.

To summarize, a comparison of the data from the three different text types suggests that significant diachronic trends and a strong preference for prominent fourth syllables are found only in narrative texts and one play (*JeuAdam*). However, a number of diachronic trends in lyric and theatrical texts approach statistical significance, and these are similar to trends attested in narrative

Pattern	r	t	$p > \|t\|$	$\bar{y}$	Slope	Constant
4only	0.461	2.986	0.005	0.069	-4.853×10^{-5}	0.086
5only	0.453	2.916	0.006	0.055	4.675×10^{-5}	0.058
2–4	0.305	1.838	0.075	0.068	-3.976×10^{-5}	0.052
2–5	0.543	3.716	0.001	0.085	9.463×10^{-5}	0.101
2–6	0.282	1.691	0.100	0.031	1.748×10^{-5}	0.034
3–5	0.509	3.394	0.002	0.112	14.529×10^{-5}	0.186
3–6	0.144	0.834	0.410	0.087	1.975×10^{-5}	0.107
4–6	0.641	4.795	0.000	0.106	-16.106×10^{-5}	0.085
2–4–6	0.421	2.663	0.012	0.110	-10.188×10^{-5}	0.108

Table 9: Regression of pattern frequency on date of composition, text type subcorpora with all texts combined

texts. Further statistical analysis is needed to establish the significance of differences between the text types.

3.2.2 Significance of differences due to text type

It appears that both the date of composition and the text type have an effect on the frequency with which different rhythmic patterns are attested. However, it must be shown that including distinctions between text types in the regression analysis provides a significantly better account of the data than a regression based on date of composition alone.

The regression analysis presented in table 9 shows all texts from the three subcorpora analysed irrespective of text type to establish the correlation between date of composition and frequency of particular patterns. The analysis shows that, taking all texts together, the date of composition is significantly correlated (at 95%) with the frequency of the majority of the patterns: a decline in the *4only*, *4–6* and *2–4–6* patterns, and a corresponding increase in the *5only*, *2–5* and *3–5* patterns. This is entirely consistent with previous findings, and illustrates that much of the variation in the rhythm of texts can be accounted for by considering differences in their date of composition alone.

To assess the statistical significance of differences according to text type, a multivariate regression analysis was used. Variables were added to the regression model, both 'dummy' variables to model the text type (e.g. [±Narrative], [±Lyric]) and interactions between these variables and the date of composition. Multivariate models were then compared to the simple model using an F-test

to assess whether they provided a significantly better model for the data.[22] The logic of the analysis is the following: if there are genuine differences in the versification of narrative as compared to lyric and theatre, then a model which takes account of the distinction between narrative and non-narrative texts should provide a significantly better fit to the data than one which only takes account of differences in date of composition.

Taking 95% as the significance level, no significant improvement on the simple regression model could be found by considering a three-way text type distinction (narrative vs. lyric vs. theatre). From figure 15, one possible reason for this is that the rhythm of theatrical texts shows features of both narrative and lyric. Like narrative, patterns with prominent fourth syllables are more common in two of the three earliest texts. Like lyric, patterns with prominent fifth syllables are relatively common at all time periods. I elected therefore to collapse the three-way distinction into two two-way distinctions, focusing first on the properties of narrative vs. non-narrative, and then on the properties of lyric vs. non-lyric texts.

If the distinction between narrative and non-narrative texts is considered, a significantly better model of the development of the *4only*, *2–4* and *2–4–6* patterns is obtained.[23] The regression coefficients are given in table 10. In all patterns, the significant (or almost significant) positive value of the narrative coefficient indicates that the pattern is consistently more common in narrative texts throughout the time period. Moreover, in the *2–4* and *2–4–6* patterns, the overall effect of date of composition (DoC) is no longer significant once the distinction between narrative and non-narrative texts is taken into account. This, and the significance of the [±Narr]×DoC coefficient in the *2–4* pattern, suggests that the overall decrease in the frequency of these two patterns is disproportionately associated with narrative texts. These results clearly show that patterns with prominent fourth syllables are most common in narrative texts. Moreover, the more iambic *2–4* and *2–4–6* patterns are only significantly

22. The method was adapted from that suggested for econometric analyses by Wooldridge (2006).

23. *2–4* pattern (including interactions with date of composition): simple model $r^2 = 0.093$, multivariate model $r^2 = 0.270$, simple model degrees of freedom = 33, multivariate model degrees of freedom = 31, comparative F-score = 3.76, $p > F(2, 31) = 0.034$. *2–4–6* pattern (including interactions with date of composition): simple model $r^2 = 0.177$, multivariate model $r^2 = 0.346$, simple model degrees of freedom = 33, multivariate model degrees of freedom = 31, comparative F-score = 4.01, $p > F(2, 31) = 0.028$. *4only* pattern (interactions not significant): simple model $r^2 = 0.213$, multivariate model $r^2 = 0.336$, simple model degrees of freedom = 33, multivariate model degrees of freedom = 32, comparative F-score = 5.97, $p > F(1, 32) = 0.020$.

Factor	Coeff.	Std. Err.	t	$p > \|t\|$
4only				
DoC	-4.010×10^{-5}	1.550×10^{-5}	-2.590	0.014
[±Narr]	0.008	0.003	2.480	0.019
Constant	0.119	0.021	5.710	0.000
2–4				
DoC	-0.227×10^{-5}	2.490×10^{-5}	-0.09	0.928
[±Narr]	0.124	0.057	2.18	0.037
[±Narr]×DoC	-8.840×10^{-5}	4.330×10^{-5}	-2.04	0.05
Constant	0.068	0.033	2.03	0.051
2–4–6				
DoC	-3.950×10^{-5}	4.390×10^{-5}	-0.9	0.375
[±Narr]	0.197	0.101	1.96	0.059
[±Narr]×DoC	-13.700×10^{-5}	7.650×10^{-5}	-1.79	0.083
Constant	0.156	0.059	2.64	0.013

Table 10: Coefficients for multivariate regression of the *4only*, *2–4* and *2–4–6* patterns on date of composition (DoC), [±Narrative] and interactions.

more common in early narrative texts; no diachronic trend is observed in lyric or theatre.

If the level at which we consider the multivariate model to be significantly better than the simple model is reduced to 90%, one further association emerges. A better model of the development of the *3–5* pattern is obtained if the distinction between lyric and non-lyric texts is considered.[24] Regression coefficients are given in table 11. The coefficients in the table suggest that while non-lyric texts show an increase in the use of the *3–5* pattern over time (positive DoC coefficient), lyric texts do not (negative [±Lyr]×DoC coefficient). Instead, the pattern is more common in lyric texts throughout the time period (positive [±Lyr] coefficient). This confirms the statistical significance of our observations from figure 15: the *3–5* pattern is more common in lyric, and lyric texts show no significant rhythmic change over time.

24. *3–5* pattern (including interactions with date of composition): simple model $r^2 = 0.259$, multivariate model $r^2 = 0.370$, simple model degrees of freedom = 33, multivariate model degrees of freedom = 31, comparative F-score = 2.73, $p > F(2, 31) = 0.081$.

Factor	Coeff.	Std. Err.	t	$p > \|t\|$
DoC	20.650×10^{-5}	4.960×10^{-5}	4.160	0.000
[±Lyr]×DoC	-19.960×10^{-5}	8.770×10^{-5}	−2.280	0.030
[±Lyr]	0.272	0.117	2.320	0.027
Constant	−0.163	0.066	−2.490	0.019

Table 11: Coefficients for multivariate regression of the *3–5* pattern on date of composition (DoC) and [±Lyric] with interaction, text type subcorpora

Pattern	$\bar{y}$	Effect		
		Overall	Narr	Lyr
3–5	0.112	↗		↑⟶
2–4–6	0.110		↘	
4–6	0.106	↘		
3–6	0.107			
2–5	0.085	↗		
4only	0.069	↘	↑	
2–4	0.068		↘	
5only	0.055	↗		
2–6	0.031	↗		

Table 12: Summary of significant effects in text type corpus

Summary

The significant effects found in the the regression analyses are summarized in table 12. Diagonal arrows indicate change over time. Vertical arrows indicate that the pattern is consistently more common in a particular text type. The horizontal arrow indicates no change over time in a particular text type. Thus, for instance, '↑⟶' for the *3–5* pattern in the lyric subcorpus shows that the pattern is consistently more common throughout in lyric texts, but does not become more significantly more common over time, unlike the effect in narrative and theatre. Patterns in the table are ranked by overall frequency in the corpus ($\bar{y}$). From table 12, it is clear that some of the diachronic trends attested in study 1 for narrative texts are attested in the wider corpus as well, in particular the decline in *4only* and *4–6* patterns, and the increase in frequency of the *5only* pattern. However, there are significant differences by text type.

In particular, narrative texts from the 12th and 13th centuries favour patterns containing prominent fourth syllables, and early narrative favours the iambic *2–4–6* and *2–4* patterns more than any other text type at any other time. By contrast, lyric texts show little rhythmic organization at any time. Plays are not significantly different from either narrative or lyric texts, suggesting that from a rhythmic point of view, they are part way between the two.

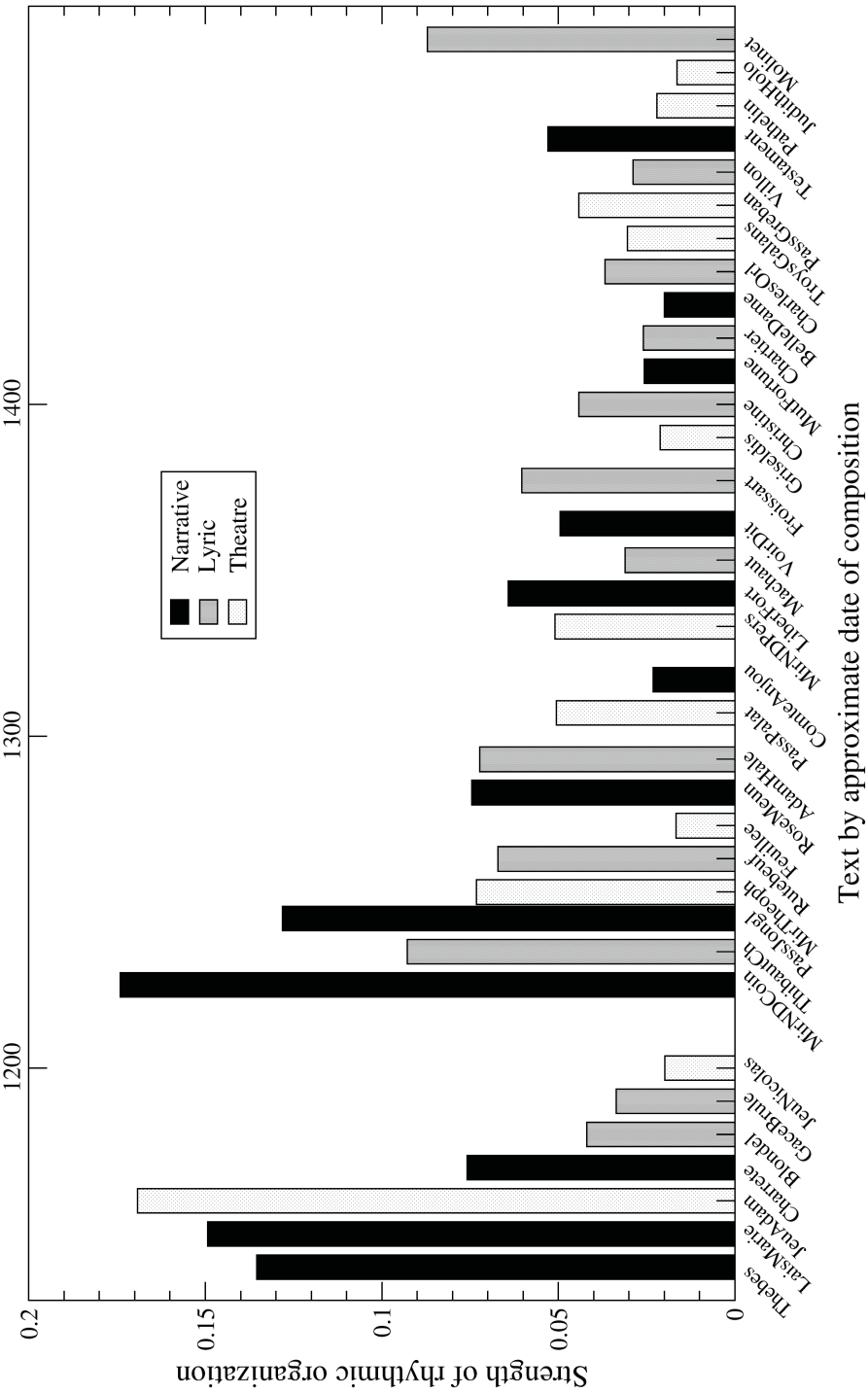


Figure 16: Strength of rhythmic organization, text type subcorpora

3.2.3 Variation between texts

Figure 16 plots the strength of rhythmic organization for the texts in the text type subcorpora. The main observation to be made from the figure is that strong rhythmic organization is largely absent from octosyllabic texts from 1250 onwards, as suggested in study 1. Notable is the strong rhythmic organization of the earliest play (*JeuAdam*); later theatrical texts rarely show strong rhythmic organization. The 'semi-dramatic' narrative *PassJongl* shows stronger rhythmic organization even than the earliest passion play (*PassPalat*). Like *BaratHaim*, Jean Bodel's *JeuNicolas* shows little rhythmic organization, lending further support to the view that the author was a metrical innovator. Finally, the strong rhythmic organization in *Molinet* is unexpected. Since the lyric extracts in the corpus contain a mixture of line lengths, the number of octosyllabic lines on which the statistics are based is smaller than for narrative and theatre, and thus there is a greater risk of studying an atypical sample. As this is an isolated finding based on a small extract, and rhythmic organization of this kind is not found in any other text from the 14$^{\text{th}}$ and 15$^{\text{th}}$ centuries, I am inclined to attribute it to a sampling error.

Comparing works of different text types by the same author, few consistent patterns emerge. For example, while the lyric verse of Guillaume de Machaut and François Villon (*Machaut*, *Villon*) is less rhythmically organized than the authors' narrative verse (*VoirDit*, *Testament*), Alain Chartier's lyric (*Chartier*) and narrative (*BelleDame*) shows a similar degree of rhythmic organization, while Christine de Pisan's lyric verse (*Christine*) is more rhythmically organized than her narrative (*MutFortune*). Adam de la Halle's *Feuillee* shows less rhythmic organization than even his lyric verse (*AdamHale*), although Rutebeuf's *MirTheoph* and lyric works (*Rutebeuf*) are more equal.

3.3 Conclusions from study 2

Study 2 has demonstrated that rhythmic organization varies little by dialect, but significantly according to text type. The strongest rhythmic organization is observed in early narrative texts, whereas lyric and theatre (with the exception of the earliest play, *JeuAdam*) never show the same degree of rhythmic organization. Guthrie (1987) too noted a difference in the rhythm of narrative and lyric verse, with lyric texts showing a less iambic rhythm than narrative. However, his suggestion that lyric verse too becomes less iambic over time finds less support in our corpus.

In chapter two (§1.1.3), I argued that the role of the musical setting in lyric performances was fundamental. The lack of organization of word stress in lyric texts is perhaps evidence that the rhythmic structure of the performance was governed by the musical setting rather than by the rhythm of the text itself. The nature of the cæsura in ten-syllable lyric texts may be taken as further evidence for this claim. Word-final schwa is permitted at the cæsura, ensuring both a regular syllable count and a regular division of the line, but causing rhythmic variation (cf. chapter two, §2.2.1). Narrative texts, on the other hand, require a regular stress at the cæsura.

Among narrative texts, I suggest that it is those most strongly linked to spoken or chanted oral performance which show the strongest rhythmic organization. In study 2, this is particularly true of the *fabliaux*, but also of religious texts such as *MirNDCoin* and *PassJongl*. In study 1, the only octosyllabic *chanson de geste*, *GormIsem*, was shown to have the strongest rhythmic organization. Although sung, *GormIsem* may clearly be classed among the *récits chantés* of Zumthor (1972), in which the simple musical setting supports the rhythm of the text itself. In romances, on the other hand, greater emphasis is placed on the written form of the text:

> La lisibilité du texte devient une valeur fondamentale : certes, même composé par l'écrit, le texte resta longtemps destiné à la déclamation ou à la lecture orale ; une transformation néanmoins s'est produite dans le mode d'audition.
>
> (Zumthor, 1972: 340)

Both Zumthor and Marnette emphasize that no absolute distinctions can be drawn between performed and non-performed texts. For example, Marnette (1998: 74) notes that Béroul's *Tristan* (last quarter of the 12th century) shows interactions between narrator and audience of the type which are elsewhere characteristic of the *chansons de geste*. Equally, in our corpus, some romances (e.g. *Eracle*) show strong rhythmic organization. The performance repertoire of jongleurs included a wide range of texts, from *chansons de geste* to *fabliaux* and romances (Duggan, 1989). Vitz (1999) goes further, claiming that romance could be composed orally and performed from memory. Nevertheless, many romances were composed in written form by clerks, and studies of narrative technique such as those of Zumthor and Marnette show fewer oral features in romance in comparison to the *chansons de geste*. The question of association with 'oral performance' is also problematic for the *fabliaux*.[25] Cailly argues that

25. In part, this difficulty stems from the relationship between the *fabliaux* and romance.

fabliaux were fundamentally oral, noting the frequency of interactions between performer and audience in the majority of texts. Orality is thus written into the text when the performed version was written down (2007: 219). Yet such interactions between author and audience are often formulaic (Cailly, 2007: 122), and it could be that these interactions are formalized features of the genre, rather than simply an indication of orality.[26] Nykrog suggests instead that the *fabliaux* were diffused in a written form, arguing that extant *fabliaux* represent the majority that were circulating in the medieval period, since there are few references in the texts to lost *fabliaux*, and some are preserved in a similar form in multiple manuscripts (1973: 40). Nevertheless, while questions of orality and genre remain controversial, it is suggestive that romances — traditionally argued to have the strongest association with written composition and diffusion, and shown in studies of discourse such as that of Marnette (1998) to have fewer marks of orality — show the weakest rhythmic organization.

Strong rhythmic organization is not attested in plays. While clearly a form of spoken oral performance, the performance context of a play is very different from that of a spoken narrative. If linguistic rhythm, simple melody and division into *laisses* serves to structure the narrative of the *chanson de geste*, the narrative of a play is instead structured through the interactions of multiple performers impersonating the characters in the story. Moreover, given that plays are told entirely through impersonation, it is possible that the lines of verse reflect a more natural spoken rhythm of the language, rather than the rhythmically stylized form inherited from the *récit chanté*. The exception is *JeuAdam*. The text was incorporated into the Latin liturgy; indeed Zumthor (1972: 443) even suggests that the dialogues in French could be considered to be elaborate liturgical responses. If so, the performance context of the text may more closely resemble that of chanted saints' lives such as *VieLeger* or *Alexis*, both of which show strong rhythmic organization.

These popular, often scatological or erotic texts were characterized by Bédier as 'la poésie des petites gens' (Bédier, 1925: 371), i.e. written for a largely illiterate, lower class public. However, scholars such as Nykrog argue instead that the *fabliaux* are written as parodies of romance, and that 'il est impossible de [les] séparer des mileux courtois' (1973: 227). Strong parallels between *fabliaux* and romances would call into question the legitimacy of separating the performance contexts of the two genres.

26. Similarly, Marnette observes that early *chansons de geste* contain fewer indications of interaction between the jongleur and his public than later *chansons*, despite the fact that the earlier texts 'sont les plus susceptibles d'avoir fait l'objet d'une récitation publique' (1998: 86).

Syllable	r	t	$p > \|t\|$	$\bar{y}$	Slope	Constant
1	0.554	2.487	0.026	0.083	-8.64×10^{-5}	0.190
2	0.394	1.604	0.131	0.227	-8.61×10^{-5}	0.334
3	0.831	5.594	0.000	0.187	40.97×10^{-5}	-0.320
4	0.839	5.772	0.000	0.475	-86.54×10^{-5}	1.546
5	0.605	2.841	0.013	0.241	21.00×10^{-5}	-0.018
6	0.047	0.176	0.863	0.246	-1.11×10^{-5}	0.260
7	0.163	0.616	0.548	0.068	1.33×10^{-5}	0.052

Table 13: Regression of frequency of PhPh edges in positions one to seven of the line on date of composition, early-8 and base subcorpus

4 Study 3: Phonological phrasing in the narrative octosyllable

In study 1, a clear evolution in the rhythm of octosyllabic verse was identified. It was argued that the change represents the loss of rhythmic organization of word stress over the course of the medieval period. This suggests in turn that stress deafness effects are unlikely to have emerged before the late 12$^{\text{th}}$ century. In this study, I examine the hypothesis that the rhythmic organization observed in study 1 is best described as a tendency to place a PhPh boundary in the middle of the line. A metrical constraint of this kind is found in ten- and twelve-syllable verse, which contain a regular cæsura. The argument advanced in study 1 is based on the assumption that the organization of word stress is salient to poets. If instead it transpires that it is the regular position of a PhPh edge which is most salient, then it is more doubtful that the strong rhythmic organization of early texts reflects a continued sensitivity to word stress. This study uses data from the early-8 and base subcorpora and PhPh edge labelling (cf. §1.2).

4.1 Phonological phrase edges in lines of verse

A first measurement of the data was carried out using the syllable method to identify the frequency of PhPh edges at each position in the line. Figure 17 shows the frequency of PhPh edges after the third, fourth and fifth syllables across the time period. A regression analysis for all positions within the line is given in table 13. The data are directly comparable to figure 3 and table 1 from

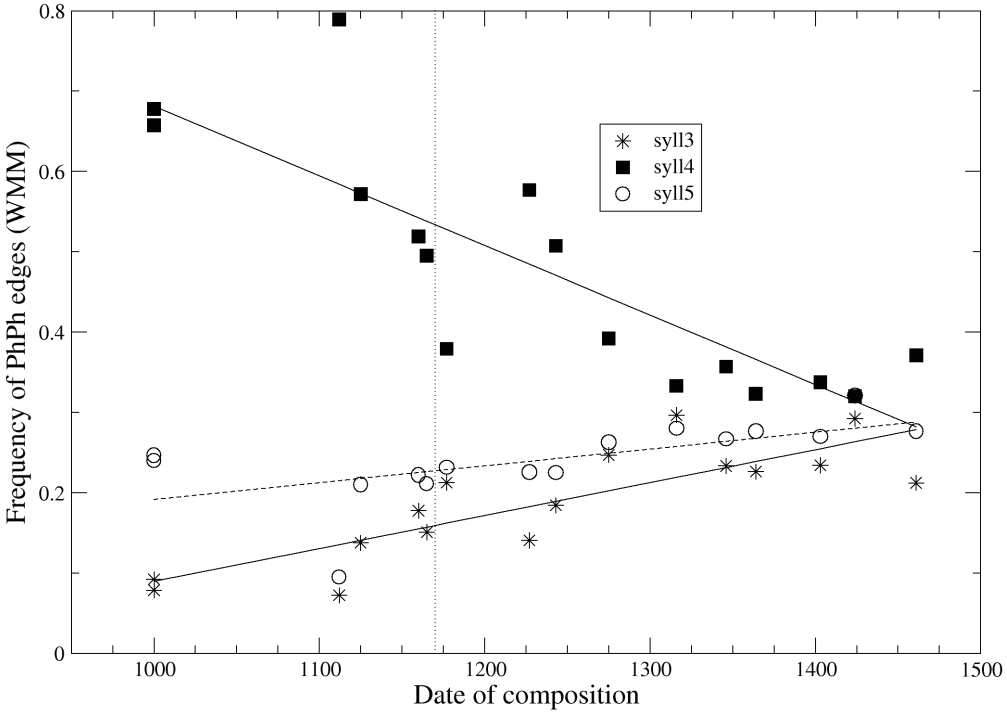


Figure 17: Frequency of PhPh edges in positions three to five of the line, early-8 and base subcorpora

study 1, which present equivalent measurements for the position of prominent syllables. The strongest trends in the data here (significant at 99%) are the decline in the frequency of PhPh edges after the fourth syllable over time, and the increase in PhPh edges after the third syllable. Nearly as strong (and still significant at 95%) is the increase in the frequency of PhPh edges after the fifth syllable over time. Taken as a whole, from figure 17, it is clear that there is a marked trend towards a loss of PhPh-based rhythmic organization. Earlier texts strongly favour a PhPh edge after the fourth syllable, but by the late 13$^{\text{th}}$ century, PhPh edges occur with broadly similar frequency after the third, fourth and fifth syllables.

These trends mirror almost exactly those shown in figure 3 in study 1. From this, we conclude that the tendency for the fourth syllable in the line to be prominent in early texts is connected to a tendency to end a PhPh in this position. Both tendencies disappear by the late 13$^{\text{th}}$ century. Therefore, the major change in the rhythm of medieval verse observed in studies 1 and 2 cannot be described purely in terms of the position of word stress, or simply

as a loss of iambic rhythmic organization.

However, there are some important differences between the data in this study and those in study 1. Firstly, there is no significant change over time in the frequency of PhPh edges after the second and sixth syllables, while study 1 showed a significant decrease in the frequency of prominent syllables in these positions. This suggests that the tendency to mark the fourth syllable with both a PhPh edge and a prominent syllable was accompanied by a weaker but genuinely word-stress based tendency to realize the second and sixth syllables as prominent. Thus, the iambic rhythm of the earlier texts is the consequence of two separate tendencies: a strong tendency to realize a prominent syllable and a PhPh edge in the middle of the line and a weaker tendency to realize a prominent syllable in the second and sixth positions.

Secondly, note that the rate at which PhPh edges after the third syllable increase in frequency over time is almost double that at which PhPh edges after the fifth syllable become more frequent (slope coefficient of 40.97×10^{-5} for syllable three vs. coefficient of 21.00×10^{-5} for syllable five). From figure 17, it is clear that PhPh edges after the fifth syllable are more common than those after the third syllable in the earliest texts. In study 1, the trends were found to be of more equal strength (39.54×10^{-5} for syllable three vs. 31.05×10^{-5} for syllable five; cf. table 1). A possible reason for this is investigated in the following section.

4.2 Mid-line mismatches

In this section, I examine the interaction between the placement of stressed syllables and the placement of PhPh edges. Where the final word in the PhPh is oxytonic, PhPh edge and stressed syllable occur at the same metrical position in the line: [27]

(30) s'es.taint le-**<u>feu</u>** / et-prant la-lan.(ce) / (*Charrete*, l. 529)

Here, the fourth syllable in the line is stressed and is followed by a PhPh edge. In common with studies of cæsura type in ten- and twelve-syllable verse (e.g. Billy, 1999), I will call this the *4m* pattern ('m' for 'masculine' ending, i.e. oxytonic).

27. In this study, I assume that the final word in a PhPh is always stressed, as in ModFr, and hence use the term 'stressed' rather than 'prominent'. This is cross-linguistically attested even in word-stress languages and is implicit, for example, in Nespor and Vogel's (1989) treatment of stress clash in northern Italian (cf. chapter four, §1.1.2).

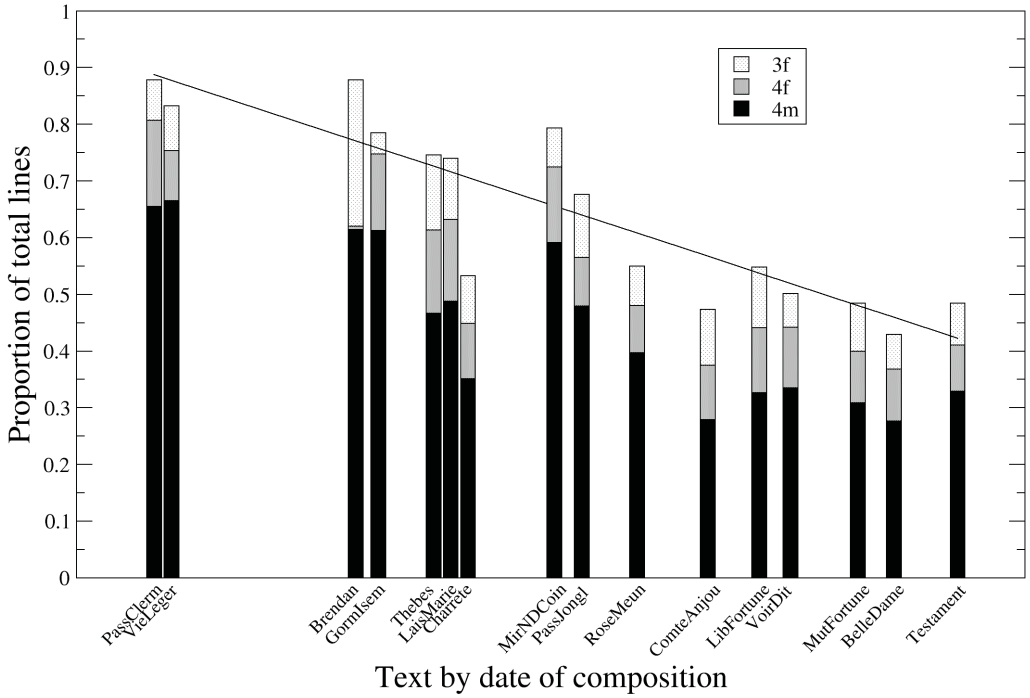


Figure 18: Marking of the fourth syllable in the early-8 and base subcorpora

However, where the PhPh ends in a paroxytonic word, the position of the stressed syllable and that of the PhPh edge are 'mismatched'. In the middle of the line, two patterns are logically possible. The first displaces the PhPh edge (*4f* pattern):

(31) a-tant lor-**ar.mes** / de.man.der.(ent) (*Charrete*, l. 583)

The second displaces the stressed syllable (*3f* pattern):

(32) ne-de-**pai.lle,** / ne-de-viez-na.(tes) / (*Charrete*, l. 513)

In meters with a regular cæsura, displacement of the PhPh edge as in the *4f* pattern is extremely rare. However, in ten-syllable lyric texts, the position of the stressed syllable can be shifted to retain the position of the PhPh edge (the lyric cæsura, cf. chapter two, §2.2.1). This is what is found in the *3f* pattern.

The relative frequency of the *4m*, *4f* and *3f* patterns was studied to establish how mismatches were treated in the octosyllable. All three patterns have in common that the fourth syllable of the line is marked prosodically, either by a PhPh-final stress (*4m* and *4f* patterns), or by a PhPh edge (*3f*, *4m*). Figure 18 shows the proportion of lines in each text in the base corpus in which one

of these mid-line marking patterns is used.[28] As we would expect, the proportion of lines with mid-line marking declines significantly over time. What is remarkable, however, is the high proportion of lines with mid-line marking in many of the earlier texts: over 80% in *PassClerm*, *VieLeger* and *Brendan*, and over 65% in every pre-1250 text except for *Charrete*. Most texts adopt a mixture of the *4f* and *3f* patterns. However, there are clearly texts which favour a stressed fourth syllable *4f* pattern over the regular placement of a PhPh edge (the *3f* pattern), notably *PassClerm*, *GormIsem* and *MirNDCoin*. The striking exception is *Brendan*, in which the *3f* pattern is used to the exclusion of the *4f* pattern. We will return to this text below.

I have assumed that the *3f* and *4f* patterns are used to mark the middle of the line. However, the *3f* pattern also marks the third syllable (with a stress), while the *4f* pattern also marks the fifth syllable (with a PhPh edge). If the assumption that these patterns are favoured because they mark the fourth syllable is correct, we predict that they will be more frequent than patterns which mark only the third or the fifth syllables:

(33) un.ne-ques.ti.**on** / leur-dis.oit (*3m*; *Thebes*, l. 171)

(34) que-le-**gué** / me-con.tre.de.ïs.(tes) (*5m*; *Charrete*, l. 791)

Figure 19 plots the ratio between the frequency of the mid-line marking patterns (*3f*, *4f*) and the corresponding patterns which mark only the third (*3m*) or the fifth (*5m*) syllable. The vast majority of pre-1250 texts (*Charrete* being the major exception) contain proportionally more instances of the mid-line marking patterns. Exceptions are caused by the infrequency of *4f* patterns in *Brendan*, and of *3f* patterns in *GormIsem*. From 1250 onwards, however, the mid-line marking patterns are not more common than the *3m* and *5m* patterns. This confirms the hypothesis that before 1250, poets are using the *3f* and *4f* patterns to mark the fourth syllable in the line, either with a stressed syllable or with a PhPh edge. The finding also explains why PhPh edges after the fifth syllable were found to be more common than those after the third syllable in earlier texts: the majority of these tokens are of the mid-line marking *4f* pattern.

28. In the remainder of this study, unlabelled syllables are assumed to form a PhPh edge. While this may be considered permissive, recall that this still discounts all word boundaries which are predicted never to form the edge of a PhPh (e.g. adjective–noun, subject pronoun–verb).

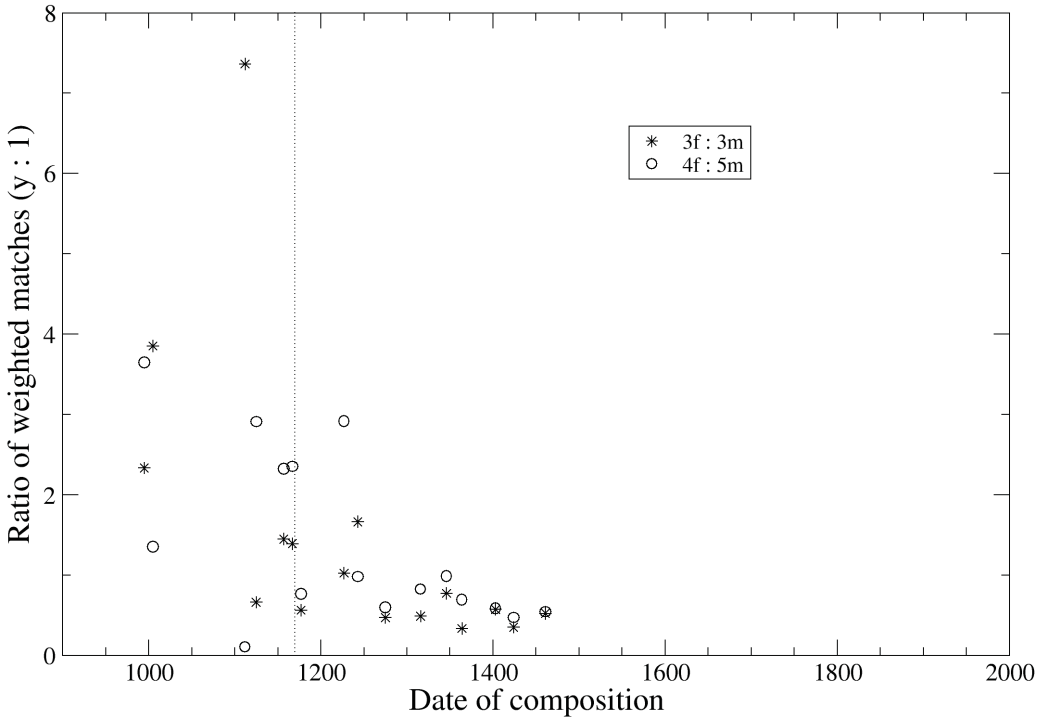


Figure 19: Ratio of stress to PhPh edge mismatched patterns to corresponding matched pattern, early-8 and base subcorpora

4.3 *GormIsem* and *Brendan*

Particularly interesting is the contrast between the two early 12$^{\text{th}}$-century Anglo-Norman texts, *GormIsem* and *Brendan*. Both show similar consistency in the use of mid-line marking, which occurs in 87.9% of lines scanned as octosyllables in *Brendan*[29] and 78.5% of lines scanned as octosyllables in *GormIsem*. Moreover, it should be recalled that this estimate only includes lines where a word boundary is also defined as a PhPh edge. Other etymologically stressed fourth syllables are not included, for example:

(35) dist lui cu.**ment-**guar.der / les-deit / (*Brendan*, l. 150)

(36) i.cos.te-**fo.le-**gent / de-Fran.(ce) / (*GormIsem*, l. 79)

(37) e-ja-l'unt **fous-**ju.eus / o.ccis / (*GormIsem*, l. 190)

Here, no PhPh edge is reconstructed after the conjunction *cument* or the pre-nominal adjectives *fole* and *fous*. However, particularly in the two *GormIsem* examples, an emphatic reading of the two pre-nominal adjectives would seem

29. Including, for this study, those with a counted final schwa syllable, cf. chapter two, §2.2.3.

plausible, suggesting a strong word stress if not a PhPh boundary in the mid-line position. Nevertheless, I am cautious in reading too much into these examples, since there are also clear cases of rhythmic variation in the poem, as will be shown below.

The strategies used for mid-line marking in the two poems are totally different. The *Brendan* poet uses a regular PhPh boundary after the fourth syllable of the line, thus alternating between the *4m* and *3f* patterns. Indeed, in *Brendan*, I could find only three tokens of lines scanned as octosyllables in which there was no word boundary between the fourth and fifth syllables, e.g.:

(38) que-plus de.**man.der** / ne-sa.vrat / (*Brendan*, l. 26)

Moreover, in only ten lines (2.16% of those that could be scanned) did an unstressed monosyllable (cf. §1.1) occur in fourth position, e.g.:

(39) pen.sez de-**la-**nef / sus trai.re! / (*Brendan*, l. 398)

I suggest that in this text, the tendency for a PhPh edge to occur after the fourth syllable is so strong that it should be considered a metrical requirement. Deviations from the meter are occasionally attested; however, this is true too of the syllable count of the poem. Only 92.2% of the lines in the extract could be scanned as octosyllables, including those with counted final schwa. These deviations may be authorial, scribal, editorial or any combination of the three, and the detailed analysis required to investigate this is beyond the scope of this thesis. However, it is clear that exceptionless metrical constraints are not attested in the text as presented by the modern editor.

The *GormIsem* poet, however, favours a regular stress on the fourth syllable of the line, alternating principally between the *4m* and *4f* patterns. While a closer look at exceptions in the *Brendan* text revealed very few cases where there was not at least a word boundary between the fourth and fifth syllables, in *GormIsem* a number of tokens show clear variation on the general rhythmic pattern. In particular, a number of formulaic expressions beginning *sil fiert* show a tendency to displace the PhPh edge to the sixth syllable, with the associated stress in fifth or sixth position:

(40) sil-fiert sur-**la-tar.ge** no.ve.(le) / (*GormIsem*, l. 50)

(41) sil-fiert sur-**sun-hel.me** ver.gié / (*GormIsem*, l. 342)

(42) si-fiert un-**che.va.lier** se.guin / (*GormIsem*, l. 454)

Other formulaic tokens showing stressed fifth syllables are also attested:

(43) a.rie.re en.**cha.ca** le-che.val / (*GormIsem*, l. 135)

The line occurs as part of a refrain repeated at the end of five *laisses* in the poem. However, variation of this kind (a *5m* or *5f* pattern) is also found in less formulaic expressions, for example:

(44) Gor.mund / li-**lan.ca** u.ne-tam.(bre) / (*GormIsem*, l. 74)

(45) al-col / sun-**e.scu** / de-quar.trés / (*GormIsem*, l. 403)

Unlike *Brendan*, I suggest that in *GormIsem* consistent mid-line marking is no more than a strong rhythmic tendency, rather than an attempt by the poet to impose a metrical constraint on the verse. Thus, lines without a regular fourth syllable stress are robustly, if not frequently, attested in *GormIsem*, while lines without a PhPh edge or at the very least a word boundary after the fourth syllable are extremely rare in *Brendan*.

The *GormIsem* octosyllable recalls to an extent the regular stress of the epic cæsura in ten-syllable *chansons de geste*. Given that the majority of early *chansons de geste* have ten-syllable lines, it is tempting to see *GormIsem* as a metrical 'experiment', retaining many of the characteristics of the longer verse form (e.g. formulaic structure, the association of clause and line, and the mid-line break), and 'expressing the same sort of epic material in lines two syllables shorter than the *usual* form' (Nichols, 1963: 507; my emphasis). Yet the dominance of the ten-syllable model in the early *chansons de geste* in the philological record may simply be due to the chance failure of octosyllabic *chansons* to be preserved. Moreover, as is clear from figure 18, the versification of *GormIsem* is not different from the earliest octosyllables, *PassClerm* and *VieLeger*. This being the case, it is far from clear that *GormIsem* reflects an atypical use of the octosyllable, susceptible to metrical influence from a 'more usual' verse form.

It is clear that mid-line marking was very common in octosyllabic verse of the early 12$^{\text{th}}$ century. However, the constraint could be interpreted in different ways. In *Brendan*, there is no evidence that the poet aimed for regular placement of stressed syllables, either in the middle or at the end of the line. Each half-line instead constitutes a prosodic unit of four syllables. In *GormIsem*, on the other hand, the position of the stressed syllables is extremely regular. The fourth syllable in the line bears a stress associated with the edge of a PhPh after the fourth (*4m*) or the fifth (*4f*) syllable in 75% of lines (cf. figure 18). *GormIsem* thus shows clear stress-based rhythmic organization, while *Brendan* does not. In study 2, texts closely associated with spoken oral performance were shown to have greater rhythmic organization. *GormIsem*, a *chanson de geste*, is written for (and perhaps even composed through) oral performance.

Brendan, on the other hand, is one of the earliest surviving verse texts with strong links to the written form, a point emphasized in the poem's dedication:

(46) *en letre* mis e en romanz [...]
de saint Brendan le bon abéth (*Brendan*, ll. 11–13)

Benedeit sees his task primarily as having recording his story 'en letres', not 'en chanson' or even 'en vers'.

4.4 Conclusions from study 3

This study has demonstrated that the loss of rhythmic organization in octosyllabic verse observed in studies 1 and 2 is connected to a decline in the frequency of mid-line marking with a regular PhPh edge. The prominent fourth syllable observed in the previous two studies frequently occurs at the end of a PhPh. The tendency to regularize the position of a mid-line PhPh boundary in the octosyllable parallels the metrical constraint on cæsura position found in ten- and twelve-syllable verse.

An examination of the treatment of 'mismatches' between the metrical position of the PhPh edge and the position of the PhPh-final stress reveals that both the pattern which retains the regular position of the stressed syllable (the *4f* pattern) and that which retains the regular position of the PhPh edge (the *3f* pattern) are used by the majority of poets. This suggests that a fixed organization of prominent syllables is not the main concern of French versifiers, which in turn casts doubt on the hypothesis that the loss of rhythmic organization in verse observed in studies 1 and 2 is connected to the loss of word stress and emergence of stress deafness.

However, while we must reject the initial conclusion that the loss of rhythmic organization in octosyllabic verse as a whole can be attributed solely to linguistic change, there is strong evidence from individual texts for the continued vitality of word stress in the 12th century and early 13th century. Texts which in study 1 showed strong rhythmic organization (*PassClerm*, *GormIsem* and *MirNDCoin*) also show a preference for the *4f* pattern in cases of mismatch, retaining the position of the stressed syllable while displacing the PhPh edge. This treatment of mismatch is elsewhere extremely rare in the French metrical tradition. With regard to the cæsura in ten-syllable verse, Duffell notes that cases of *césure enjambante*, in which an extra post-tonic schwa is added to the first hemistich causing the second hemistich to be reduced by one syllable, is 'a principle so alien to French metrics that it was termed *coupe italienne*' (2008:

86). Both Duffell (1996: 212–13) and Elwert (1965: §26) explicitly attribute this difference between French and Italian cæsural practice to the weakness of word stress in French. Yet it is precisely this 'alien' tendency to regularize the position of stress at the cost of a regular division of the line into PhPhs which we observe in some pre-1250 octosyllabic texts. Moreover, these texts were also shown in study 1 to favour iambic rhythmic patterns of prominent syllables (*2–4–6*), and while the prominent fourth syllable is linked to a PhPh edge, there is no corresponding preference for a PhPh edge after the second and sixth syllables. The strong tendency to place a stressed syllable associated with a PhPh edge in fourth position is thus accompanied by a lesser tendency to stress the second and sixth positions, creating an overall iambic rhythm. The following passage from *GormIsem*, provided for illustration, shows a sequence of lines which conform almost exactly this pattern:

(47) par.mi-le-flanc / l'e.spie li-mist /
la-bo.ne-en.sei.ne / qu'il-tint / (non-octosyllabic)
de-l'au.tre-part / en-fist eis.sir /
le-sanc ver.meil / en-fist sai.llir /
e-dist Er.nout: / "e.stez mei ci" /

(*GormIsem*, ll. 170–74)

5 Study 4: Prose controls for verse data

The final study in this chapter aims to show that patterns identified in octosyllabic verse texts in the three previous studies are indeed a conscious manipulation of linguistic patterns on the part of poets, and not simply caused by the natural rhythms or syntax of the language. Texts from the prose subcorpus were used to form lines of 'pseudo-verse', and these pseudo-lines were analysed using the same techniques as applied to real verse in previous studies. In all tests, the hypothesis is that no organization, either of prominent syllables or PhPh edges, is found in the pseudo-verse lines.

I begin by giving a brief description of how pseudo-verse is generated (§5.1), before comparing the rhythm (§5.2) and phonological phrasing (§5.3) of these pseudo-lines to that of verse.

5.1 Pseudo-verse

5.1.1 Theoretical background

The most commonly used type of prose control for verse studies is referred to by Gasparov (1987) as the 'speech model'. The speech model works by analysing sections of prose which by chance also conform to the meter of verse. Such a technique is used for English verse by Tarlinskaja and Teterina (1974) and for MedFr verse by Noyer (2002), and is also used in this study. In each case, pseudo-verse is used to determine the chance frequency of iambic rhythm occurring in the language, and is then compared with verse texts in order to test whether the poet has included a greater than chance frequency of iambic lines. The verse text compared to the speech model by Tarlinskaja and Teterina (1974) is John Donne's *Satyres*, which, while it does not correspond to the normal constraints of English iambic verse, is shown to contain more iambic lines than chance frequency, leading the authors to conclude that it is to some extent iambic. Noyer (2002) includes pseudo-verse data from a single prose text (Villehardouin's *Conqueste de Constantinople*) in order to show that all 12$^{\text{th}}$-century octosyllabic verse texts in his corpus show a greater than chance frequency of iambic organization.

Speech models, or pseudo-verse, do not provide perfect controls. Gasparov (1987: 323) notes in particular that the speech model may be biased by the rhythm of the prose text. Noyer's (2002) use of a prose control is problematic, as the text dates from *c.*1209, some 30 years later than the latest verse text to which it is compared. In both Noyer's study and the present investigation, where diachronic change is so important to the argument, a single control text of this kind is insufficient, and the current study makes use of a wider sample.

5.1.2 Creating pseudo-verse

Pseudo-verse for the current study was automatically generated from the texts in the prose subcorpus using the labelled PhPh edges. For the purposes of pseudo-verse construction, I assume that lines of verse must be delimited by PhPh boundaries.

Octosyllabic pseudo-verse definition

A sequence of eight metrical syllables, of which the final metrical syllable is a full vowel (`v`). The sequence may also contain an uncounted final schwa syllable. The sequence must be preceded and followed by a PhPh edge.

Only those syllables labelled as PhPh edges were treated as valid line breaks; unlabelled syllables were not. In defining 'metrical syllables', elision of post-tonic schwa before a vowel was consistently observed, and in 15th-century texts, elision of word-internal schwa in hiatus with a following vowel was also obligatory (cf. chapter two, §2.2.1).

From the prose extracts of approximately 3,000 words, between 118 (*VieLouis*) and 196 (*AdvisionChr*) octosyllabic pseudo-lines were generated. Below is an example of three pseudo-lines drawn from a single sentence in *AdvisionChr*:

(48) La m'apparoit l'estature / d'un-homme / de-belle-forme /
mais-de-grandeur inextimable, / car-sa-teste / tresperçoit les-nues, /
ses-piez / marchoient les-abismes / et-son-ventre / avironnoit
toute-la-terre. / (*AdvisionChr*, I, §1, ll. 15–18)

(49) mais de gran.deur in.ex.ti.ma.(ble)

(50) ses piez mar.choi.ent les a.bis.(mes)

(51) a.vi.ro.nnoit tou.te la ter.(re)

While pseudo-lines do not form a continuous text, each in isolation is a plausible line of verse.

5.2 Rhythm of pseudo-verse

In study 1, it was concluded that the development of octosyllabic verse showed a loss of rhythmic organization over the course of the 13th century. If this is a change in versification, rather than a linguistic change, it should be possible to show that pseudo-verse from the 12th and 13th centuries is no more rhythmically organized than pseudo-verse from the 14th and 15th centuries. Moreover, if as claimed in study 1 verse from the 14th and 15th centuries shows no rhythmic organization, it should show a similar variety of rhythmic patterns to pseudo-verse extracts.

The position of prominent syllables in the pseudo-verse lines was measured using the syllable method, and a selection of stress curves are shown in figure 20. While the stress curves are fairly flat, there remains a hint of a preference for stress on even positions in the earlier texts: positions two and four in *QuatrLivre* and positions four and six in *TristanPr*. A χ^2 test on the hypothesis that the position of prominent syllables in positions two to six of the pseudo-line is not random failed to produce significant results for ten of the eleven extracts (the exception being the 15th-century *QuadrInvec*). Despite hints at iambicity

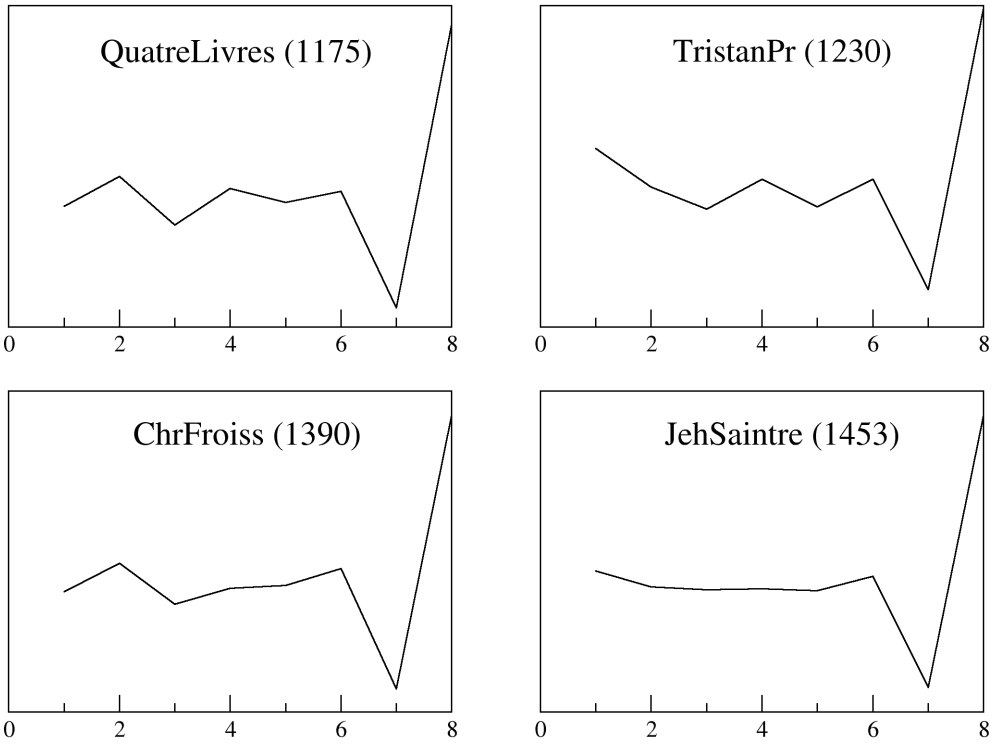


Figure 20: Stress curves for selected prose texts, octosyllabic pseudo-verse

in the earlier texts — which if shown to be significant over a larger pseudo-verse sample may indicate that versification of the 12th and 13th centuries exaggerates a linguistically common rhythmic pattern — it is clear that the far stronger rhythmic organization of genuine verse from this period represents a manipulation of linguistic rhythm by the poet.

The only pseudo-verse extract to show significant rhythmic organization is the 15th-century *QuadrInvec*, and the stress curve for this text is given in figure 21. The curve shows that prominent syllables are most common in the fifth position of the pseudo-line. The stress curve suggests that *QuadrInvec* contains a high proportion of PhPhs ending in a PWd of three syllables, with positions six, seven and eight of the line containing a single word, such that the previous word ends in fifth position. In study 1, it was found that *BrevNobles*, also by Alain Chartier, strongly favoured a 3+3 organization of the second hemistich: again, this suggests a preference for trisyllabic PhPh-final PWds. Moreover, the two most common rhythmic patterns in Chartier's octosyllabic *BelleDame* text are the *2–5* and *3–5* patterns (cf. data points for 1425 on

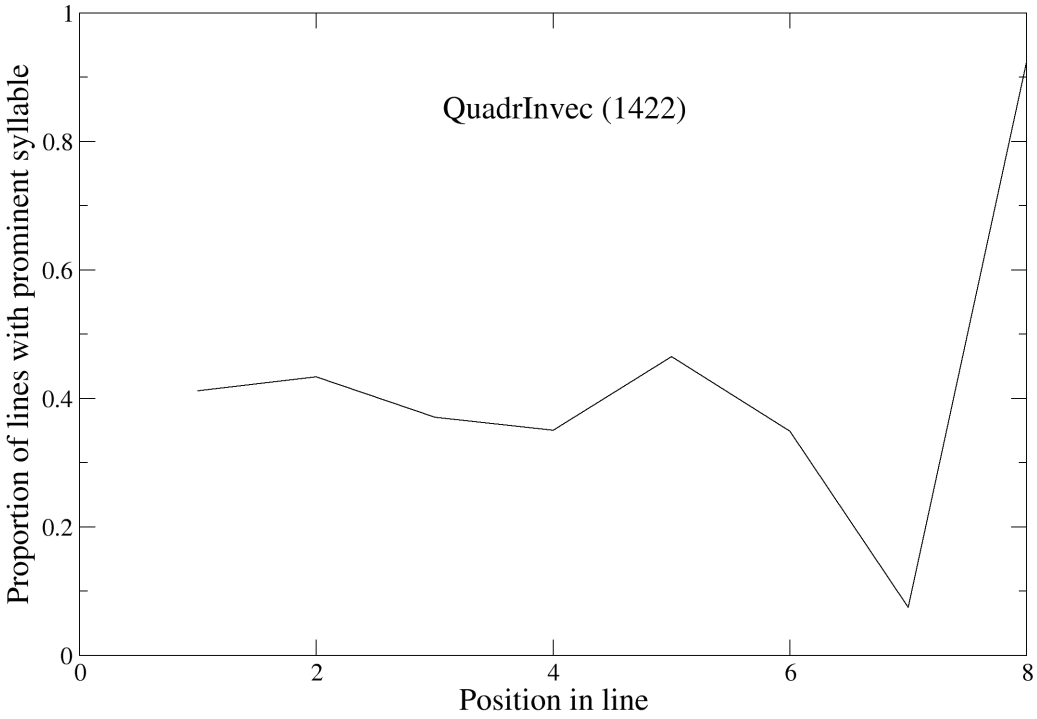


Figure 21: Stress curve *QuadrInvec*, octosyllabic pseudo-verse

figure 6). In our view, these data do not suggest that Chartier aimed for mid-hemistich division in *BrevNobles* as suggested in study 1, but that he used a disproportionately high number of trisyllabic PWds across all his works. The increased frequency of prominent fifth syllables in octosyllabic verse and the continued predominance of prominent third syllables in six-syllable hemistichs in the 15$^{\text{th}}$ century may suggest that trisyllabic PWds at the end of the PhPh became more common in the language as a whole. This may reflect the influx of longer words borrowed from Latin in the 14$^{\text{th}}$ and 15$^{\text{th}}$ centuries. However, the trend appears to be particularly strong in Chartier's work.

From 1250 onwards, I have argued that rhythmic organization was lost from verse. To show that verse of this period is rhythmically indistinguishable from prose, a regression analysis was carried out to test the hypothesis that the proportion of lines with a prominent syllable in a particular position is correlated with the factor [±Pseudo]. The analysis is shown in table 14. The only significant difference (at 95%) is found in the proportion of (pseudo-)lines with prominent fourth syllables. The weighted match mean for strong fourth syllables is on average 0.043 greater in genuine verse from 1250–1500 than

Syllable	r	t	$p > \|t\|$	$\bar{y}$	Slope	Constant
2	0.380	0.860	0.406	0.375	0.017	0.384
3	0.267	1.428	0.177	0.334	0.019	0.338
4	0.376	2.383	0.033	0.413	−0.043	0.430
5	0.416	1.510	0.155	0.368	0.030	0.363
6	0.397	0.808	0.433	0.423	−0.014	0.425

Table 14: Regression of frequency of prominent syllables in positions two to six of the (pseudo-)line on [±Pseudo], base and prose (pseudo-verse) subcorpora from 1250–1500

in pseudo-verse, a difference of approximately 10%. From these data, we can conclude that verse of the 14th and 15th centuries shows very little rhythmic organization, and none that cannot be explained by a slight tendency to retain the mid-line stress so common in verse of the 12th century. The weakness of rhythmic organization in later French verse is demonstrated by Gasparov (1987), who uses a similar comparison of verse and prose to study the classical alexandrine (twelve-syllable line). He concludes that 'the stress structure of both hemistiches in the French *alexandrin* is determined only by the rhythmical vocabulary of the French language and nothing else; there are no additional constraints in arranging stresses in the line' (1987: 334).

5.3 Mid-line marking in pseudo-verse

It was shown in study 3 that one of the most important changes in the octosyllable was the loss of a mid-line PhPh boundary, a change which also affected the position of prominent syllables in the line. It was assumed in study 3 that the use of a mid-line break was a conscious feature of the verse form, and its loss simply represents a change in versification practice. An examination of pseudo-verse confirms this assumption. There is no significant change in the proportion of pseudo-lines with a PhPh edge after the third, fourth or fifth position over time (contrast figure 17 for genuine octosyllabic verse). In no sample of pseudo-verse are the mismatched mid-line marking patterns *3f* and *4f* more common than the *3m* and *5m* patterns (contrast figure 19).

However, there is evidence from the pseudo-lines that the strong tendency to mark the mid-point of an octosyllabic line may have its origins in the rhythm of the language. Taking an average over all the texts, figure 22 shows the most

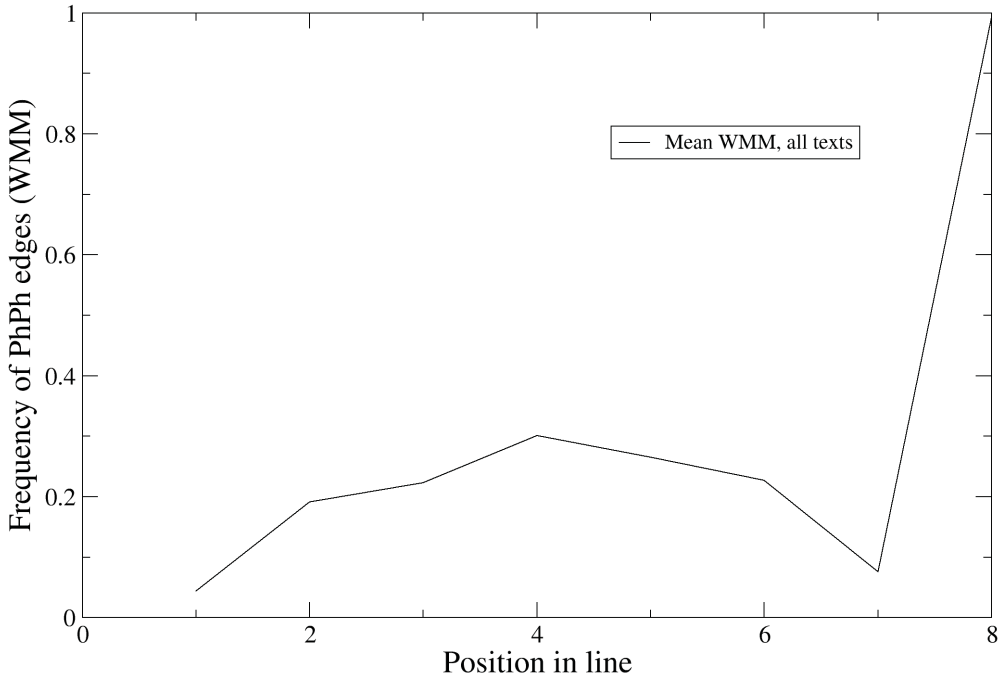


Figure 22: Position of PhPh edges in lines of octosyllabic pseudo-verse, prose subcorpus

common position of PhPh edges overall in the octosyllabic pseudo-line. PhPh edges in the first and seventh positions in the pseudo-line imply a monosyllabic PhPh. Even though rhythmic factors were not taken into account in the labelling, the rarity of PhPh edges in these positions in figure 22 shows that monosyllabic PhPhs are extremely uncommon in MedFr. Moreover, the figure suggests that PhPh edges after the fourth syllable of the pseudo-line are marginally more common than in other positions. This may reflect a tendency to produce PhPhs of equal length, dividing the eight-syllable group of the pseudo line into two groups of four syllables. [30]

A further piece of evidence which hints at a correspondance between verse structure and linguistic structure is found if the pseudo-texts are considered individually. A χ^2 test on the hypothesis that PhPh edges are not randomly distributed in positions two to six of the pseudo-line demonstrates that in *Quatr-*

30. Such a constaint has been suggested for the construction of ModFr stress groups. For example, Delais-Roussarie (1995) argues for a EQUI constraint: 'au sein d'un constituant C^{i+1}, chaque constituant prosodique contient le même nombre de syllabes' (cited in Lacheret-Dujour and Beaugendre, 1999: 150).

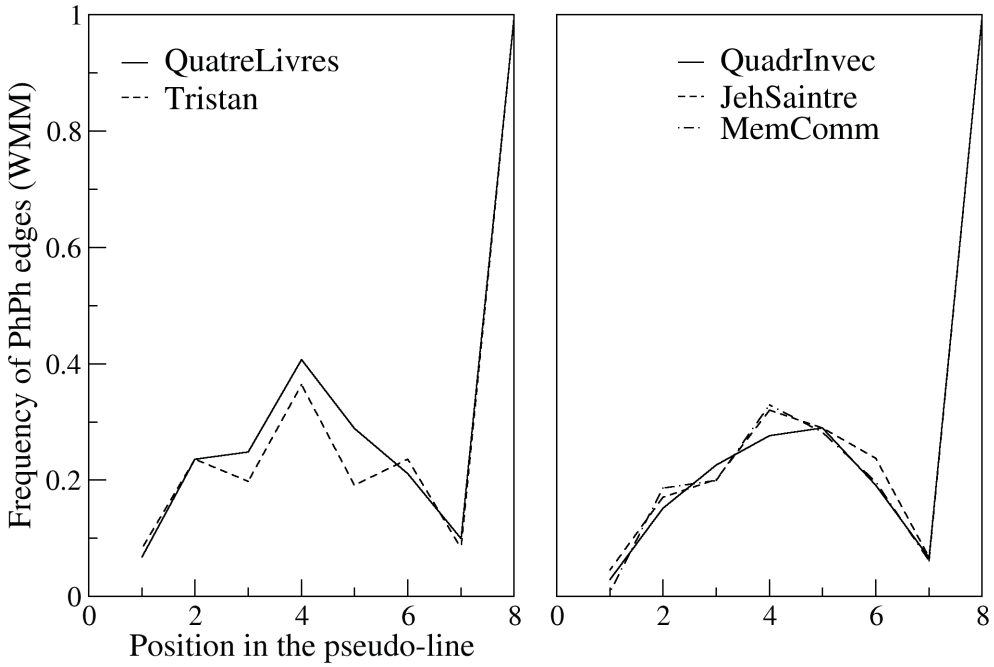


Figure 23: Position of PhPh edges in octosyllabic pseudo-verse, extracts showing non-random positioning in syllables two to six.

Livre, *TristanPr*, *QuadrInvec*, *JehSaintre* and *MemCommyn*, PhPh edges are more common in certain positions than in others. [31] Given the highly artificial nature of pseudo-verse (only a part of the whole prose extract is studied), I do not consider it likely that this finding represents an important difference between texts. Two of the texts showing non-random positioning of PhPh edges are from before 1250; the remainder are from the 15th century. Figure 23 plots a curve showing the frequency of PhPh edges in each position in the pseudo-line in these texts, overlaying the two pre-1250 texts in the left-hand plot and the three 15th-century texts in the right-hand plot. In the two earlier texts, 'non-random' positioning of PhPh edges is caused by a frequent division of the pseudo-line into two PhPhs at the fourth syllable. In the 15th-century texts, however, 'non-random' positioning is caused by a high frequency of PhPh edges

31. *QuatrLivre*: $\chi^2 = 13.78$, $p = 0.008$. *TristanPr*: $\chi^2 = 11.51$, $p = 0.021$. *QuadrInvec*: $\chi^2 = 11.06$, $p = 0.026$. *JehSaintre*: $\chi^2 = 10.16$, $p = 0.038$. *MemCommyn*: $\chi^2 = 12.77$, $p = 0.012$. All with 4 degrees of freedom.

in both fourth and fifth position. This difference shows striking parallels to the rhythmic changes attested in genuine verse in study 1, and the changes in the position of PhPh edges attested in study 3. The fact that similar trends are attested in pseudo-verse suggests that one factor contributing to rhythmic change in MedFr verse is a change in the rhythmic structure of the language itself, in particular the average length of the PhPh. It must be stressed that this suggestion is much more speculative than conclusions reached elsewhere in the chapter. No statistically significant diachronic trend can be demonstrated across all pseudo-verse extracts, and the pseudo-verse itself is only a subset of the prose data. Nevertheless, it would be incorrect to assume that the rhythm of French is constant throughout the medieval period.

5.4 Conclusions from study 4

Study 4 has established that the rhythmic organization detected in verse texts of the 12th and 13th centuries is a feature of the verse form, and not caused by the chance occurrence of such patterns in the language. However, there is some evidence that the rhythm of versification represents an exaggeration of slight rhythmic tendencies in the language, such as the tendency to divide an eight-syllable prosodic constituent into two four-syllable halves in the 12th and 13th centuries. Verse of the 14th and 15th centuries, except for having a slightly higher than chance tendency to contain a prominent fourth syllable in a line, does not show any other properties that differentiate it greatly from the rhythmic patterns found in contemporary prose texts. In other words, its meter is purely syllabic.

Chapter summary

The studies in this chapter had two overall aims: to demonstrate the existence of the organization of word stress in the earliest texts and, if successful, to establish a timeline for its disappearance. The results from study 1 are very persuasive on both counts, at least as regards the octosyllable. There is strong evidence for rhythmic organization in octosyllabic texts from the 11th and 12th centuries, favouring stress on even syllables and particularly on the fourth syllable. Such rhythmic organization disappears by the mid-13th century. In study 4, it was shown that the rhythm of eight-syllable 'pseudo-lines' drawn from prose texts is almost indistinguishable from that of verse in the

14th and 15th centuries, conclusively demonstrating that rhythmic organization had been lost.

However, two factors emerged from the other studies which call into question the hypothesis that this change may be linked directly to the emergence of group stress and stress deafness. In study 2, the versification of octosyllabic narrative was found to be substantially different to that of lyric texts and plays. Moreover, a more detailed examination of narrative reveals that those texts with the closest links to spoken oral performance show the strongest rhythmic organization. In one sense, this is a positive finding, as it is precisely in perfomed texts without the support of a complex musical setting that we may except to see linguistic rhythm manipulated to structure the text. However, in another sense this finding is problematic, since later verse texts do not show the same links to spoken oral performance: *fabliaux* are not attested beyond the end of the 13th century, and religious narratives such as *MirNDCoin* and *PassJongl* are increasingly presented as plays (e.g. *MirNDPers* and *PassPalat* from the 14th century). Verse narratives of the 14th and 15th centuries owe more to the tradition of the romance and the lyric poem than to the tradition of the *récit chanté*. Thus, the disappearance of rhythmic organization from verse could be explained purely as a change in the context of performance of verse texts.

In study 3, it was shown that much of the rhythmic organization of the octosyllable attested in study 1 results from a tendency to end a PhPh after the fourth syllable in the line. This suggests that the rhythmic organization of the octosyllable may not have depended on the position of word stress, but rather on the position of PhPh-final stress, a principle attested in ten- and twelve-syllable verse and clearly compatible with a group-stress grammar.

However, in *PassClerm*, *GormIsem* and *MirNDCoin* the edge of the PhPh tends to be shifted to ensure that the primary stress of a paroxytone is realized on the fourth syllable. This prioritization of the placement of stress over a regular syllabic division of the line is not attested in any verse form in French beyond the early 13th century. Clearly, the placement of stressed syllables remained important to these authors, and despite the evident perceptual saliency of the PhPh edge at this time, I suggest that this type of versification is not compatible with either stress deafness or the loss of word stress. Consequently, while the evidence is considerably more complex than suggested in study 1, there remains evidence from versification to date the emergence of group stress to the late 12th century or even the early 13th century.

Chapter 4

From word to phrase: Reanalysis of stress and its wider effects

In chapter three, we examined evidence from verse texts to investigate the chronology of the emergence of group stress, and concluded that the earliest date at which such a system could have developed was the late 12th century. This chapter returns to the two remaining problems identified in the introduction: the mechanism by which change in the stress rules took place, and whether it had any wider consequences.

In section 1, I will consider the mechanism of change directly. From a reconstruction of common stress patterns in the noun phrase using both word- and group-stress grammars, I will assess whether there was sufficient overlap between the predictions of the two possible grammars in the 12th and 13th centuries to permit a reanalysis to take place. The consequences of prosodic change form the main focus of sections 2 and 3, which return to two issues raised in chapter one (§3.2). In section 2, I will examine whether the emergence of stressed post-verbal pronouns constitutes evidence for the existence of a group-stress grammar. In section 3, I will discuss the effect of the emergence of group stress on the verb-second word order of MedFr, and will investigate the hypothesis that a loss of stress on the initial constituent contributed to the decline of verb-second in the 14th and 15th centuries.

1 The reanalysis of stress

The study in this section examines the hypothesis that the stress system of MedFr, particularly in the pre-1250 period, could be reanalysed as a group-stress system. For this to occur, the stress pattern produced by adult speakers with an underlying word-stress grammar must be sufficiently ambiguous for a child learner to develop a group-stress analysis of the data.

In this study, I begin by examining the stress patterns of six types of PhPh containing a head noun and one other PWd. The predicted output of a word-stress grammar for each of the six types of PhPh is then compared to the stress pattern predicted by a group-stress grammar to evaluate the degree of overlap between the two. Using data from the syntactically tagged subcorpora (early-8, base and prose), I then consider the relative frequency of the different types of PhPh to establish the proportion of PhPhs in MedFr which are incompatible with a group-stress analysis.

1.1 Background

1.1.1 The domain of the phonological phrase

In chapter 3 (§1.2), a method was outlined for labelling the edges of PhPhs in MedFr, based purely on the syntactic structure. A definitive reconstruction was shown not to be possible, as the syntactic theory assumed by the prosodic phonology models of the 1980s (Nespor and Vogel, 1986; Selkirk, 1986) was not clearly applicable to the clause structure of MedFr. The clearest theoretical predictions related to the noun phrase, an area in which the syntax of MedFr more closely resembles that of modern Romance languages. Moreover, Post's (2000) corpus-based study of the ModFr PhPh, which assesses the validity of the theory, relates exclusively to the noun phrase. For these reasons, the present study will focus only on PhPhs headed by a noun (henceforth N-PhPh).

The N-PhPh consists of a head noun and all preceding words within the noun phrase, whether function or content words. Additionally, the optional rule of PhPh restructuring predicts that noun–adjective sequences may in some cases form a PhPh (cf. chapter one, §1.1.2; chapter three, §1.2). In the present study, I will initially treat all noun–adjective sequences as a single PhPh in MedFr. However, in section 1.2.3, I will consider the patterns of PWds favoured by different syntactic combinations. If noun–adjective sequences are shown to be atypical in their prosody, this assumption may have to be revised.

1.1.2 Word stress and clash resolution

The second stage of the analysis is to reconstruct the stress pattern of the N-PhPh as predicted by a word-stress grammar. The principal difficulty here is the treatment of stress clash, for example:

(1) je ne pris pas [**plain poing**]$_{\text{PhPh}}$ de cendre (*Charrete*, l. 798)

Stress clash of this kind is cross-linguistically uncommon without some form of resolution (cf. chapter one, §1.1.1). Nespor and Vogel (1989) suggest two possible strategies for clash resolution, exemplified initially with northern Italian data. Firstly, the first of the clashing stresses may be deleted (beat deletion):[1]

(2) [sarà me**tá** strada]$_{\text{PhPh}}$ (Nespor and Vogel, 1989: 76)

In (2), the stress on the oxytonic adjective *metá* is deleted, due to a clash with the following stress on *strada*. Secondly, the perceptual distance between the clashing stresses may be increased, either through pause insertion or through the lengthening of the first stressed vowel (beat insertion) (1989: 79). In this case, neither stress is deleted. In northern Italian, this strategy is used only where PhPh-final stresses clash:

(3) [la veritá]$_{\text{PhPh}}$ [vince]$_{\text{PhPh}}$ quasi sempre (*ibid.*: 79)

In (3), both the stress on *veritá* and the initial stress on *vince* are retained, with the authors noting lengthening or pause insertion between the stressed syllables.

If MedFr behaves like modern northern Italian, then we predict that the first of the clashing word stresses in (1) will be deleted. Nespor and Vogel (1989) point out, however, that the level on the prosodic hierarchy at which beat deletion and beat insertion apply varies from language to language. In Greek, for example, stress deletion only occurs within the PWd (following incorporation of extra morphemes), while clashing stress between PWds is resolved by beat insertion (Nespor and Vogel, 1989: 91–92). The only other Romance example discussed by the authors is modern Catalan, which is argued to behave identically to northern Italian with regard to clash resolution (1989: 89–90). Given that both Catalan and northern Italian are closely related to French within the Romance family, I will assume that clash resolution by

1. More specifically, the authors predict that the first of the two clashing stresses is deleted if the second stress is final in the PhPh (1989: 76–77). We will only be considering PhPhs containing two PWds in this section, so the second word will always be PhPh-final.

deletion of word stress was also a feature of MedFr. [2]

Lastly, Nespor and Vogel (1989) argue that clash resolution is an accent deletion rule. In chapter one, we instead assumed that clash resolution was a stress retraction rule, causing stress to be realized earlier in the first word if it has more than one syllable (cf. also Liberman and Prince, 1977; Nespor and Vogel, 1979; Hayes, 1984, 1995). However, the two positions can be reconciled. The deletion analysis finds support from acoustic phonetic studies. In English, Horne (1990b) and Vogel et al. (1995) both argue that in stress clash contexts the fundamental frequency of the clashing syllable is greatly reduced. However, no concomitant rise in fundamental frequency on the first syllable is attested (Horne, 1990b: 978; although Shattuck-Hufnagel, 1991 disagrees). Post (2000: 64) reaches the same conclusion for ModFr. 'Shifted' stress is argued to be a perceptual, rather than a phonetic effect (Vogel et al., 1995: 124; see also Grabe and Warren, 1995). Thus, although a pitch accent is not present in the speech signal, I assume that a stress shift is perceived in clash resolution contexts, and will continue to mark it as such in reconstructions of MedFr prosody.

1.1.3 The prosodic word

In a word-stress grammar, every PWd is assumed to be stressed. However, not all words are PWds. In chapter three (§1.1), a number of common function words (prepositions *a*, *de*, weak object pronouns, determiners) showing a historically unstressed development were argued to belong to a category of unstressed (or type 1) monosyllables. I assume that these are not PWds in MedFr, but prosodic clitics in the sense of Selkirk (1996).

Selkirk (1996) proposes three structures by which prosodic clitics may be incorporated into the prosodic hierarchy:

Free clitic: [fnc (lex)$_{\text{PWd}}$]$_{\text{PhPh}}$ [3]

Internal clitic: [(fnc lex)$_{\text{PWd}}$]$_{\text{PhPh}}$

Affixal clitic: [(fnc (lex)$_{\text{PWd}}$)$_{\text{PWd}}$]$_{\text{PhPh}}$ (Selkirk, 1996: 188)

Free clitics are not PWds, but nor are they incorporated into a PWd: they are immediately dominated by the PhPh. This is the analysis of ModFr function words adopted by Post (2000: 85), and it is the analysis I adopt for MedFr. The internal clitic analysis must be ruled out, since this incorrectly predicts

2. Clash resolution has also been proposed for ModFr (Mazzola, 1992; Hoskins, 1994; Post, 2000); however, such an analysis implies that there is an underlying word-level stress clash to be 'resolved'. I have instead argued that word stress is irrelevant in ModFr, cf. chapter one, §1.

3. 'Fnc' = function word, 'lex' = content word.

PWds	Pre-1250 (%)	Post-1250 (%)
1	61.2	60.1
2	33.6	32.0
More	5.2	7.9
Total	12 texts, 6076 tokens	15 texts, 8025 tokens

Table 1: Number of PWds in the N-PhPh, early-8, base and prose subcorpora, all texts

that only word-internal phonological rules may apply between clitic and host. For example, the deletion of schwa before a following vowel is regular across word boundaries from the earliest texts; however, the deletion of schwa before a vowel within a word is not attested before the 14th century (cf. chapter two, §2.1, §2.2.2). Since prosodic clitics such as *le* and *la* are regularly elided, a PWd boundary must exist between the clitic and the following word. The affixal clitic analysis too is unlikely: in ModFr, it can be ruled out by the observation that group-initial secondary stress does not fall on initial clitics (cf. chapter one, §1.3). Post (2000: 82–83) accounts for this by arguing that the clitic is not part of a PWd, contrary to the predictions of the internal or affixal analyses.

1.2 The stress pattern of the N-PhPh

The corpus analysis is presented in two parts. Firstly, I investigate the frequency of N-PhPhs containing one, two, or more than two PWds. Secondly, focusing only on those N-PhPhs containing two PWds, I examine the stress pattern and frequency of six different combinations of PWds in the corpus, in order to establish the proportion of cases in which the word-stress and group-stress grammars produce compatible stress patterns.

1.2.1 Number of prosodic words

The first analysis of the corpus data identifies the number of N-PhPhs containing one, two, or more PWds. The data are presented, divided into pre- and post-1250 texts, in table 1. Dividing the corpus into two halves chrono-

logically, the overall distribution of the three classes of N-PhPh was found to be significantly different in the two time periods.[4]

The majority of tokens contain only a single PWd, the head noun. Here, modulo the presence of prosodic clitics, the PWd and the PhPh are co-extensive. These tokens thus cannot be used to distinguish word- and group-stress analyses, and will henceforth be disregarded.

Only a small proportion of tokens contain more than two PWds. The data show a significant correlation between the proportion of these PhPhs and the date of composition of the text, with later texts showing more such tokens.[5] The texts showing the highest rates of these N-PhPhs are prose texts from the 14$^{\text{th}}$ and 15$^{\text{th}}$ centuries: *AdvisionChr* (11.3%), *MirLouis* (10.9%), *JehSaintre* (9.7%), *ChrFroiss* (9.4%), *MemCommyn* (9.3%). The verse vs. prose distinction is not statistically significant once date of composition is taken into account, and the verse text showing the highest proportion of these N-PhPhs is also the latest (*Testament*, 9.1%). Three of the prose texts with high rates of more than two PWd N-PhPhs contain a number of tokens which are characteristic of written prose style. For example, proper names are often accompanied by a title, a demonstrative, and an epithet: e.g. *li dis messires Hues*, *chils biaux roys Phelippes* (*ChrFroiss*), and there are a number of tokens of *du benoiet saint loys* from *MirLouis*. These legalistic, over-specific noun phrases are particular to this kind of prose 'witness' account, and are unlikely to have been present in the spoken data to which acquirers of French were exposed. N-PhPhs containing more than two PWds represent only a small proportion of the total, and there are many more possible combinations of different PWd lengths, making the study of possible stress patterns complex. For this reason, the remainder of the study will focus on N-PhPhs with only two PWds.

1.2.2 Stress patterns of two prosodic word N-PhPhs

The most important factor in determining the stress pattern of a N-PhPh using a word-stress grammar is the relative length of the PWds. In particular, monosyllabic PWds are likely to trigger a stress clash. Six principal structures will be considered:

1m#1 [σ]$_{\text{PWd}}$ [σ (ə)]$_{\text{PWd}}$
treis feiz; *de bon mestre*; *a grant peine*

4. $\chi^2 = 40.7$, $p = 0.000$ (2 degrees of freedom).
5. $r = 0.499$, $t = 2.88$, $p > |t| = 0.008$

1f#1 [σ ə]$_{\text{PWd}}$ [σ (ə)]$_{\text{PWd}}$
telle peine; *la droite veie*

1#n [σ (ə)]$_{\text{PWd}}$ [σ (...) σ (ə)]$_{\text{PWd}}$
a grant honor; *la bonne valee*

nm#1 [σ (...) σ]$_{\text{PWd}}$ [σ (ə)]$_{\text{PWd}}$
de commun cours; *son second frere*

nf#1 [σ (...) σ ə]$_{\text{PWd}}$ [σ (ə)]$_{\text{PWd}}$
joyeuses choses

n#n [σ (...) σ (ə)]$_{\text{PWd}}$ [σ (...) σ (ə)]$_{\text{PWd}}$
par commune renommee; *un piteux entremes*

In the definitions, constituents in parentheses are optional, thus (ə) signifies with or without final schwa, (...) signifies with or without intervening syllables. In the labels, '1' signifies 'one full syllable', 'n' signifies 'at least two full syllables', and where necessary 'f' and 'm' specify whether the word ends with a post-tonic schwa syllable (f) or not (m).

Neither type *1#n* nor type *n#n* show stress clash, thus stress is located on the final non-schwa syllable of each PWd:

1#n [σ (ə)]$_{\text{PWd}}$ [σ (...) σ (ə)]$_{\text{PWd}}$
a grant honor; *la bonne valee*

n#n [σ (...) σ̲ (ə)] [σ (...) σ̲ (ə)]
par commune renommee; *un piteux entremes*

Where the second word is monosyllabic and the first word is oxytonic (*1m#1* and *nm#1*), there is a stress clash. Assuming clash resolution takes place, the stress pattern of these sequences is predicted to be as follows:

1m#1 [σ] [σ̲ (ə)]
treis feiz, *de bon mestre*; *a grant peine*

nm#1 [σ̲ (...) σ] [σ̲ (ə)]$_{\text{PWd}}$
de commun cours; *son second frere*

In the *1m#1* sequence, stress on the first PWd is deleted. In the *nm#1* sequence, stress on the first PWd is perceived to be shifted on to the initial syllable.

For the remaining two types (*1f#1* and *nf#1*), stress clash is blocked only by a post-tonic schwa, and the following patterns result:

1f#1 [σ̲ ə] [σ̲ (ə)]
telle peine; *la droite veie*

nf#1 [σ (...) σ ə] [σ (ə)]
joyeuses choses

In the analysis that follows, I assume that schwa is sufficient to prevent a stress clash. However, it seems plausible that stress clash may have taken place in spite of the presence of schwa. Jacobs (1992: 68) proposes that schwa is phonologically extrametrical. Extrametrical constituents are 'invisible for the purposes of [stress] rule application' (Hayes, 1995: 57), and this is clearly true of MedFr schwa, which is invisible to the regular word stress rule. If extrametrical for the purposes of stress assignment, it is also possible that schwa is invisible for the application of clash resolution. If this is the case, the stress pattern of type *1f#1* would be identical to that of type *1m#1*, with initial stress deleted, while type *nf#1* would show an identical stress pattern to type *nm#1*.

With the predictions of a word-stress grammar clarified for the six types of two-PWd PhPhs, we can now predict which types could be produced using a group-stress grammar without a change in the surface stress pattern. In the ModFr group-stress grammar, primary stress is regularly PhPh-final, and secondary stress is realized on the first syllable of the first content word, provided this is not adjacent to the phrase-final stress (cf. chapter one, §1) Below, the predictions of word- and group-stress grammars for each type of PhPh are given side by side:

	Word Stress	Group Stress
1m#1	[σ] [σ (ə)]	[σ] [σ (ə)]
1f#1	[σ ə] [σ (ə)]	[σ ə] [σ (ə)]
1#n	[σ (ə)] [σ (...) σ (ə)]	[σ (ə)] [σ (...) σ (ə)]
nm#1	[σ (...) σ] [σ (ə)]	[σ (...) σ] [σ (ə)]
nf#1	[σ (...) σ ə] [σ (ə)]	[σ (...) σ ə] [σ (ə)]
n#n	[σ (...) σ (ə)] [σ (...) σ (ə)]	[σ (...) σ (ə)] [σ (...) σ (ə)]

Only in the *nf#1* and *n#n* types do the two grammars predict different stress patterns. Therefore, there appears to be considerable potential for reanalysis.

1.2.3 Frequency of the six types of N-PhPh

Table 2 shows the frequency with which each of the six combinations of PWds are attested in the corpus. Patterns *1m#1* and *1f#1*, along with *nm#1* and *nf#1*, are combined at the end of the table, in keeping with the suggestion that the presence of a post-tonic schwa in the *1f#1* and *nf#1* patterns may have had no effect on clash resolution.

Pattern	Pre-1250 (%)	Post-1250 (%)	Verse (%)	Prose (%)	Overall (%)
1m#1	21.0	16.4	21.4	14.8	18.4
1f#1	14.6	10.7	14.0	10.6	12.5
1#n	51.6	53.4	52.9	52.2	52.6
nm#1	4.1	4.0	3.5	4.6	4.0
nf#1	1.6	2.4	1.7	2.3	2.0
n#n	7.2	13.2	6.4	15.4	10.5
Total	100	100	100	100	100
1#1	35.6	27.1	35.4	25.5	30.9
n#1	5.6	6.3	5.3	6.9	6.0

Table 2: Proportion of prosodic patterns in the N-PhPh, early-8, base and prose subcorpora

The majority of two-PWd N-PhPhs show the *1#n* pattern, in which clash resolution does not apply and both word- and group-stress grammars generate identical outputs. The only cases in which word- and group-stress grammars make different predictions, the *n#n* and (possibly) *nf#1* patterns, are found in just over 10% of tokens overall. The proportion is lower still in texts from before 1250. The *nm#1* pattern, where clash resolution causes stress to be shifted, is relatively rare throughout. However the *1m#1* pattern, in which clash resolution *deletes* the stress from the first PWd, is common, accounting for 21.0% of all two-PWd N-PhPhs in texts from before 1250.

Differences according to date of composition and verse vs. prose are significant for the *1m#1* and *n#n* patterns. *1m#1* tokens become less frequent over time, while *n#n* tokens are more frequent in later texts.[6] Even taking into account the fact that the corpus contains more early texts in verse than in prose, the difference is still significant: *1m#1* patterns are more common in verse than prose throughout the time period, while *n#n* patterns are especially common in later prose texts.[7]

6. Regression of *1m#1* on date of composition: $r = 0.498$, $t = -2.87$, $p > |t| = 0.008$. Regression of *n#n* on date of composition: $r = 0.474$, $t = 2.69$, $p > |t| = 0.013$.

7. *1m#1*: regression on date of composition only, $r^2 = 0.248$, regression on date of composition and [±Verse], $r^2 = 0.420$, F(1, 24) = 7.147, $p = 0.013$ (two variable model significantly better). Date of composition coefficient = -13.3×10^{-5} ($p = 0.064$), [±Verse] coefficient = 0.050 ($p = 0.013$), constant = 0.332 ($p = 0.002$). *n#n*: regression on date of composition only, $r^2 = 0.224$, regression on date of composition, [±Verse] and interaction, $r^2 = 0.511$, F(2, 23) = 6.752, $p = 0.005$ (three variable model significantly better). Date of

Pattern	Overall (%)
A N	33.0
N A	6.7
N N	9.9
Fnc N	34.7
Fnc cl N*	15.7
Total	100

* Clitic intervenes between words, e.g. *sulunc la lei*, *devant la porte*.

Table 3: Syntactic structure of two-PWd N-PhPhs, overall averages, early-8, base and prose subcorpora

The texts with the highest proportion of *n#n* tokens in the post-1250 period are *QuadrInvec* (22.5%), *AdvisionChr* (34.9%) and *JehSaintre* (19.3%), and all prose texts from *Berinus* onwards show 11.0% or above of such tokens. The highest proportion found in a verse text is 10.8% (*Testament*). In the 15th-century prose texts, there is frequent use of polysyllabic adjectives in both pre-nominal and post-nominal position: e.g. *diverses passions*, *merveilleuse advision*, *pardurable arrousement*, *de jugement apostolique*, *de perdicion singuliere* (*AdvisionChr*); *en perilleuses aventures*, *mauvaises ambitions*, *proufitable sentence*, *des voluptez delicieuses*, *de punicion divine*, (*QuadrInvec*); *des amoreux desirs*, *de la couronne muralle* (*JehSaintre*). In *JehSaintre*, many tokens are of *le petit Saintré*, which may account for the high rate of *n#n* tokens in this text. One linguistic change which is likely to have contributed to this increase is the influx of Latinate vocabulary in the 14th and 15th centuries. Of the examples cited above, *advision*, *apostolique*, *ambitions*, *voluptez* and *muralle* are all borrowed from Latin, and are not attested in 12th-century texts.[8] However, there are also substantial differences between individual texts: the poetic prose style of Christine de Pisan and Alain Chartier leads to a greater number of tokens in *AdvisionChr* and *QuadrInvec* than any other text. Indeed, *AdvisionChr* shows over twice the proportion of *n#n* N-PhPhs as that found in the later *MemCommyn* text.

composition coefficient $= 42.5 \times 10^{-5}$ ($p = 0.012$), [±Verse] coefficient $= 0.427$ ($p = 0.089$), [±Verse]×DoC coefficient $= -37.8 \times 10^{-5}$ ($p = 0.049$), constant $= -0.423$ ($p = 0.056$).

8. Source: *Trésor de la langue française*, <`http://atilf.atilf.fr/`> [accessed 21 Jan-

Pattern	*1m#1* (%)	*1f#n* (%)	*n#n* (%)	Other (%)	Total (%)
A N	19.1	15.3	11.9	53.7	100
N A	9.4	11.0	27.1	52.6	100
N N	17.5	3.9	14.0	64.5	100
Fnc N	27.1	18.0	3.4	51.4	100
Fnc cl N	2.2	0.1	13.9	83.8	100

Table 4: Syntactic structure of two-PWd N-PhPhs by prosodic pattern, early-8, base and prose subcorpora

Table 3 shows the breakdown of two-PWd N-PhPhs by syntactic structure. Table 4 shows the proportion of these tokens with patterns *1m#1*, *1f#1* and *n#n* marked. This contrasts the most common pattern in which word- and group-stress produce different outputs (*n#n*) with the most common pattern causing stress on the first word to be deleted (*1m#1*).

Overall, the most common two-word combinations of PWds in the N-PhPh are adjective–noun and function word–noun. Only a small proportion of these tokens show the *n#n* pattern which distinguishes the word- and group-stress grammars (3.4% of function word–noun combinations and 11.9% of adjective–noun combinations). Thus, the overwhelming majority of tokens of adjective–noun, noun–noun and function word–noun combinations show a stress pattern which is compatible with a group-stress grammar.

Moreover, a larger proportion of adjective–noun, noun–noun and particularly function word–noun combinations show the *1m#1* pattern, where the stress on the first PWd is deleted. This shows that even with a word-stress grammar, 19.1% of all pre-nominal adjectives were unstressed due to clash resolution, as were 27.1% of all function words. The extensive deaccenting of PWds, I suggest, favoured the emergence of the group-stress analysis.

In noun–adjective combinations, which constitute only 6.7% of two-PWd N-PhPhs, the *n#n* pattern is much more frequent, accounting for 27.1% of tokens, while the *1m#1* pattern is much rarer, accounting for only 9.4% of tokens. This suggests that noun–adjective sequences are less likely to be reanalysed as having a group stress. In ModFr, this is exactly what is found. Post (2000) argues that only monosyllabic post-nominal adjectives are incorporated into the ModFr PhPh (cf. chapter one, §1.1.2). Thus, noun–adjective *n#n* patterns

uary 2011].

form two separate PhPhs in ModFr, retaining the final stress on the noun.

To summarize the findings, in the majority of N-PhPhs in MedFr, word- and group-stress grammars are predicted to generate identical stress patterns. In a small minority of cases, the innovative group-stress grammar is incompatible with the output of the word-stress grammar. In noun–adjective combinations, where such incompatible tokens are proportionally more common, no reanalysis takes place, and the noun and adjective form separate PhPhs in ModFr. In other syntactic combinations, tokens incompatible with group-stress are outnumbered by those showing deletion of word stress through clash resolution. Moreover, tokens of stress clash are more frequent in the earliest texts, while tokens incompatible with a group-stress analysis are rare. This supports the hypothesis that group stress is most likely to have emerged in the earlier part of the time period studied, i.e. before 1250.

1.3 Summary

In this section, I have shown that both word-stress and group-stress grammars produce identical stress patterns for the majority of N-PhPhs in MedFr. The assumption that obligatory clash resolution applied within the MedFr PhPh as it does in other Romance varieties is crucial to the analysis. Within a word-stress grammar, clash resolution is predicted to have resulted in the frequent deletion of PhPh-internal stresses, leading either to stress not being realized at all (if the word was stress-initial), or being weakly realized on the initial syllable of the word (if the word was not stress-initial). Moreover, the proportion of tokens showing stress clash is higher in the earlier texts. We are now in a position to answer the first question that this chapter set out to address: from the evidence of the N-PhPh, the potential for reanalysis of the stress rules existed in the 12$^{\text{th}}$ century.

The results of this analysis leave a number of questions open for future research. Firstly, only the N-PhPh was considered. The findings presented here should be compared with other common phonological phrase groupings (e.g. those containing the finite verb) to investigate whether or not the same patterns of PWds are attested. However, as the domain of the PhPh containing the finite verb is less clearly defined (cf. chapter three, §1.2), this analysis would not be simple to carry out.

A second question is at what point the proportion of tokens distinguishing the two grammars becomes large enough to prevent reanalysis. With the exception of noun–adjective combinations, 10–15% of tokens of two-PWd N-

PhPhs are predicted to have been incompatible with a group-stress grammar, yet this did not prevent reanalysis of the stress system as a whole. In the case of noun–adjective combinations, 27.1% of tokens of two-PWd N-PhPhs distinguished word- and group-stress grammars. Here, stress on the first PWd is still realized in ModFr.

2 The development of post-verbal non-subject pronouns

In chapter one (§3.2), we considered the theory that a fixed phrase-initial stress in MedFr affected word order. One of the phenomena frequently attributed to such a prosodic rule is the Tobler-Mussafia law: unstressed non-subject pronouns cannot occur in clause-initial position in any early Romance language (Tobler, 1875; Mussafia, 1886; more recent treatments in Benincà, 1995, 2006). In practice, the Tobler-Mussafia law applies to verb-first constructions, where the lack of a pre-verbal constituent places the verb group in clause-initial position. In such cases, at least until the late 12$^{\text{th}}$ century, non-subject pronouns are realized post-verbally.

In this section, I will focus on one of the most interesting aspects of this verb–pronoun inversion: the use of stressed forms (e.g. the strong forms *moi*, *toi* and *soi*) in post-verbal position in French. Not only does this pose problems for the prosodic account of the Tobler-Mussafia law (Ramsden, 1963: 124; and chapter one, §3.2.2), but as it is unique to French of all Romance languages, it may reflect unique features of the stress system. [9]

In this section, because the development in question begins early in the history of French, I will examine the whole of the *Vie de saint Alexis*, in addition to all extracts in the full corpus. [10]

9. I assume that the post-verbal positioning of pronouns is due to syntactic verb movement (cf. chapter one, note 44). Following Cardinaletti and Roberts (1991), Benincà (1995, 2006), I assume that post-verbal pronouns are not syntactically enclitic to the finite verb, and thus cannot be incorporated into the same PWd (cf. Selkirk, 1996: 205).

10. The three earliest texts were also examined, although no tokens of interest to the study were found.

2.1 Background

2.1.1 Three types of post-verbal pronoun

In ModFr, all post-verbal pronouns bear PhPh-final stress.[11] For the purposes of the following analysis, post-verbal forms are divided into three types. Firstly, the post-verbal form may be an allomorph of the pre-verbal form (i.e. a strong form instead of a weak form):[12]

(4) donne-**moi**

This type includes the pronouns *moi*, *toi* and *soi*, and I will refer to it henceforth as the MOI type. Secondly, the post-verbal form may be a stressed realization of the pre-verbal form, and otherwise morphologically and phonologically identical (i.e. a post-verbal weak form):

(5) vas-**y**

(6) dites-**nous**[13]

(7) prenez-**les**

Also included in this type are pronouns *en*, *la*, *vous*, and MedFr *li*, and I will refer it as the Y-NOUS-LES type. Thirdly, there is the form *le*. This post-verbal weak form is the only instance in ModFr in which an etymological schwa is fully stressed, adopting the allophone [ø]:

(8) prends-**le** (realized [lø])

2.1.2 Historical lack of stress

Post-verbal stressed forms are not a regular etymological development. In other Romance languages, post-verbal pronouns remain unstressed. Compare ModFr and modern standard Italian:

(9) dis-**moi** (ModFr)

(10) dim**mi** (standard Italian)

In Italian, the etymological stress on the finite verb is retained, and the weak post-verbal pronoun is unstressed.

Moreover, there is some limited evidence from the earliest French texts that post-verbal pronouns are historically unstressed. In *PassClerm* and *VieLeger*,

11. Unless followed in the PhPh by a short adverb, e.g. *aidez-moi donc* (de Kok, 1985: 293).

12. While the distinction between strong and weak forms is historically stress-based, by the textual period the distinction is purely morphological, cf. chapter one, §3.2.2.

13. On *nous* as a weak form, cf. chapter one, fn. 41.

there are three tokens in which the consonantal form of a pronoun is used before a following consonant.[14] Two tokens are of the MOI type:

(11) que de sa mort **posche·s** neger. (*PassClerm*, l. 238)

(12) por ciel tiel duol **rova·s** clergier (*VieLeger*, l. 65)

One case of reduction of *le* to *l* is attested:

(13) et cum il l'aut doit de ciel'art,
rende·l qui lui lo comandat (*VieLeger*, ll. 25–26)

This list is exhaustive: no other such tokens of post-verbal enclitic consonantal forms exist in any of the earliest French texts. Slightly more robust evidence for the etymologically unstressed nature of post-verbal pronouns stems from cases in which a consonantal form is used before a following vowel. There are two such tokens of *s* (reflexive) in *VieLeger*, and two of *l*, e.g.:

(14) **torne s'**als altres, si lor dist (*VieLeger*, l. 206)

(15) cum si l'aut fait, **mis l'**en reclus (*VieLeger*, l. 155)

Storey's (1968) edition of the *Vie de saint Alexis* contains one token of *m*, one of *s* and one of *l* before a vowel. However, of these only the token of *l* is present in the manuscript; the remaining two tokens are editorial corrections based on the syllable count of the line:

(16) cil vait, sil quert, **fait l'**el muster venir (*Vie de Saint Alexis*, l. 181)

(17) venent devant, **jetent s'** [**MS sei**] an ureisuns (*ibid.*, l. 357)

(18) **aidiez m'** [**MS mei**] a plaindra le deul de mun ami (*ibid.*, l. 462)

While example (16) is uncontroversial, it is difficult to interpret examples (17) and (18). On the one hand, given that the base manuscript dates from the early 12th century (thus perhaps half a century after the text was written) and is preserved in a different dialectal form, it is quite plausible that the scribe could have eliminated weak post-verbal consonantal pronouns, which to him were ungrammatical. On the other hand, despite the editor's emendations, the syllable count of the poem is not entirely regular, and there remain forty of the 625 lines with an incorrect count even after elision has been manipulated (Storey, 1968: 30–31). Moreover, elsewhere in the poem, strong forms are used post-verbally:

14. There is one additional token from *PassClerm*, in which the editor (but not the manuscript) suggests reduction of *les* to *ls*:

(i) e per es mund **roa·ls** [**MS l**] allar
toz babtizar in trinitad (*PassClerm*, ll. 453–54)

(19) **quer mei**, bel frere, ed enca e parcamin (*ibid.*, l. 281)

The evidence from these 11th-century texts is very limited, and requires careful interpretation. From *PassClerm* and *VieLeger* we see that consonantal pronouns *s* (reflexive) and *l* could be used post-verbally. Moreover, it is not until the *Vie de saint Alexis* that the first tokens of strong pronoun forms in post-verbal position are attested. [15] For the purpose of the present study, we retain the hypothesis that post-verbal pronouns are historically unstressed, and moreover are still unstressed in the early 11th century. Subsequently, post-verbal pronouns become stress-bearing. In the MOI type, this has a morphological effect: the weak form is replaced by its strong allomorph. In the Y-NOUS-LES type, there is no morphological effect, but weak forms become stress-bearing, as in ModFr. *Le* will be considered separately because of its schwa vowel.

2.1.3 Phonological phrasing

Since non-subject pronouns are consistently realized in pre-verbal position as prosodic clitics, I assume that they must be incorporated into the PhPh of a neighbouring lexical head. Consequently, there are two possible configurations for phonological phrasing. Firstly, the pronoun may be incorporated into a following PhPh:

(I) $[\ \text{V}\ ]_{\text{PhPh}}\ [\ \text{Prn}\ (...)\ ]_{\text{PhPh}}$ (non-restructured)

Alternatively, PhPh restructuring may take place, with the pronoun incorporated into the PhPh of the finite verb:

(II) $[\ \text{V Prn}\ ]_{\text{PhPh}'}\ [\ (...)\ ]_{\text{PhPh}}$ (restructured)

In this case, the pronoun is the last element in the PhPh.

Where the pronoun is proclitic on a following vowel (e.g. 16 above), I take this as clear evidence for phrasing I. On the other hand, where no material follows the pronoun within the clause (e.g. 19 above), only phrasing II is possible. I claim that stressed post-verbal pronouns only emerge under phrasing II. Whether this is the result of a group-stress grammar or not is an issue to which I return in section 2.3.

15. In *PassClerm*, strong forms are written in the same way as unelided weak forms (e.g. *per me*, l. 262; *de me·t membres*, l. 295), so no clear conclusions can be drawn from examples such as *fai se revivere* (l. 35): *se* could be weak or strong, stressed or unstressed.

2.2 The situation from the 12th century onwards

2.2.1 The MOI type

From a historical point of view, the development of the MOI type is the easiest to trace since the emergence of stress-bearing post-verbal pronouns causes an orthographically distinct strong form to appear in post-verbal position. These are first attested in *Alexis*, and in the corpus are regularly used in post-verbal position in the 12th century:

(20) entendez **moi** // nobile chevalier (*CharNimes*, l. 337)

(21) serviront **toi** // .III.M. conpaignon (*CharNimes*, l. 310)

(22) deslie **moi** cestui prison;
claimme **moi** quite sa prison (*Charrete*, ll. 916–17)

In the examples above, it is unlikely that scribal intervention of the type potentially seen in *Alexis* has occurred. In (21) and (22), while the manuscript dates from the late 13th century, the form occurs at the epic cæsura, a position with a regular metrical stress. Example (22) is chosen for its philological soundness: it is from *Charrete*, preserved in a manuscript from the early 13th century by a known Champenois scribe.

There are a number of cases in which strong forms are not used. Firstly, Foulet (1924) notes that weak forms are used before the subject pronoun in the 13th century:

(23) ha Dex, **fait se il**, pere nostre!
(*Roman de Renart*, branch XII,[16] l. 214; Foulet, 1924: 56)

Before a nominal subject, however, a strong form must be used:

(24) **fet soi** Renart, esta un prestres
(*ibid.*, branch I,[17] l. 821; Foulet, 1924: 56)

Foulet's explanation of this alternation is that the subject pronoun is included in the stress group of the finite verb, but that the nominal subject is not (1924: 58). Thus, in (24), the non-subject pronoun is final in the stress group, while in (23) it is not. In our terms, both (23) and (24) show phrasing II, but in (23), the subject pronoun is also included in the restructured PhPh.

The second exception is dialectal. In Picard, Foulet (1924) argues that the strong disjunctive forms *mi ti si*, found after prepositions, can never be used in post-verbal position. Weak *me te se* forms are used instead:

16. *c.*1202, ed. Martin (1882–87).
17. Last quarter of the 12th century.

(25) ho Walet biaus niés, va **te** sir.
pour Dieu, sire, voeilliés **m'**oïr
(*Feuillee*, ll. 363–64; Foulet, 1924: 65)

Foulet (1924: 66) maintains that the weak form in *va te sir* is fully stressed. Attested cases of elision, such as in *voeilliés m'oïr*, must therefore represent a 'forced usage' by poets. In our terms, Foulet assumes that phrasing II applies in Picard as well, but that the MOI type develops like the Y-NOUS-LES type, with stress on a weak form.

An alternative explanation is that Picard shows phrasing I. We assumed in chapter three (§1.2) that the elision of monosyllables is evidence that the monosyllable is in the same PhPh as the following word. Foulet's analysis of *voeillés m'oïr* in example (25) is therefore problematic. However, the example is far from isolated. In the Picard texts in the corpus, elision of weak MOI-type post-verbal pronouns is the rule rather than the exception:

(26) me.tés **me͡ en** bui.es et en fers (*ComtePoit*, l. 169)

(27) biaus maistres, consillié **m' [MS me]** aussi (*Feuillee*, l. 246)

(28) donne **m' [MS me]** assés de pois pilés (*Feuillee*, l. 343)

(29) laissié **m' [MS me]** aler car je sui rois (*Feuillee*, l. 395)

However, it is notable that all manuscripts (although not Langlois's (1951) edition of *Feuillee*) show the full weak form of the pronoun rather than the consonantal form. Nevertheless, the syllable count of these tokens consistently gives an elided reading. Indeed, only one non-elided token is found:

(30) giete **moi** a honor de ci (*Eracle*, l. 5730)

I discuss the use of the Francian strong form *moi* below.

Where no material follows the post-verbal pronoun in the clause, phrasing I is predicted to be impossible. In our corpus, no tokens of weak MOI forms were found in these constructions. Instead, strong forms in *-oi* and *-ei* are attested:

(31) "beau sire, se deus vos aït,
hebergiez **moi** par charité,
si ferez aumosne et bonté!" (*BouchAbev*, l. 60–62)

(32) que vous avés dites le **moi**
et je vous en aidrai en foi (*ComtePoit*, ll. 237–38)

In example (31), I read *par charité* as dislocated, leaving *moi* in clause-final position. In example (32), the rhyming of *moi* and *foi* strongly suggests that

the use of the *moi* form is authorial rather than scribal. Moreover, strong forms in *-oi* are attested in contexts where phrasing I would also be possible:

(33) dame donés **moi** vostre amor (*ComtePoit*, l. 180)

(34) donés **moi** tant de repentance (*ChvBarisel*, l. 829)

Weak forms are therefore only possible in Picard texts in contexts where phrasing I is possible. Strong forms in *-oi* are attested, although not as consistently as in other dialects, and are obligatory in phrasing II contexts. This leads me to suggest, contra Foulet, that weak post-verbal MOI forms in Picard are not stressed.

The third exception to Foulet's generalization which emerges from our corpus is lexical. Weak forms are common post-verbally in a small number of expressions. For example, as late as the 14th century, two weak post-verbal forms are attested in *MirNDPers*:

(35) vez **me** cy (*MirNDPers*, l. 341)

(36) laissier **m'**ester (*MirNDPers*, l. 194)

Tokens of *laisser m'* plus infinitive (36) are especially common: there is a token of *laissié me aler* (elision required by syllable count) in *Feuillee*; two of *lessiez me en pais* (elision again required), one in *MirTheoph* and one in *Behaingne*; and a further *lessez m'ester* token in *LaisMarie*. Both exceptions (35–36) are compatible with a phrasing I analysis.

From the analysis of the MOI type of post-verbal pronouns, we conclude that strong forms are regular in post-verbal position by the 12th century, and exceptionless in contexts where only phrasing II is possible. In cases where phrasing I is possible, Picard texts commonly retain weak post-verbal forms into the 13th century. In other dialects, however, the alternation between weak forms before a following pronoun and strong forms before any other constituent strongly suggests that phrasing II has generalized, except in a small number of relatively fixed expressions.

2.2.2 The Y-NOUS-LES type

The pronouns of the Y-NOUS-LES type do not have equivalent strong forms (cf. chapter one, §3.2) or consonantal forms, and so orthography and syllable count are of no use in determining their phrasing. In fact, the only way of determining whether pronouns of this type were stressed or not is to consider their use at the obligatorily stressed rhyme and cæsura.

Pronouns of the Y-NOUS-LES type are first attested at the rhyme in 12th-century texts:

(37) ceste foiz moustrez **le nous**
(*Roman de Thebes*, l. 10104; Skårup, 1975: 402)

(38) mais se tu as nul autre ami
n'amie nule, envoie **m'i,**
qui n'ait de moi mal faire envie.
(Chrétien de Troyes, *Perceval*,[18] l. 2288; Skårup, 1975: 391)

(39) des qu'il vos plest, ralons **nos an**
(Chrétien de Troyes, *Chevalier de la Charrete*, l. 1995; Skårup, 1975: 384)

(40) al samadi lur vient uns mes,
de la part Deu salüet **les**.
(*Brendan*, ll. 405–6; also Skårup, 1975: 402)

Tokens are rare and few are present in our corpus, so I have drawn from the wealth of examples of post-verbal pronouns in Skårup (1975). Of these, example (40) with *les* at the rhyme is both unique in our corpus and the only example I could find in the secondary literature of a direct object pronoun at the rhyme. Indeed, tokens are so rare that in his extensive study of enclitic pronouns, Melander (1928: 99) could find none. It is also the only piece of evidence to my knowledge to show that a /les/ pronunciation (rather than /ləs/) for the plural direct object pronoun was available by the 12th century.[19]

All the examples in the previous paragraph are clause-final, and so phrasing II is obligatory. However, two tokens of the Y-NOUS-LES type at the epic cæsura demonstrate that stressed forms were also used in non clause-final position, where phrasing I would also be possible:

(41) succurrat **nos** // li reis od sun barnét (*ChRoland*, l. 1061)

(42) asoldrai **voz** // pur voz anmes guarir (*ChRoland*, l. 1133)

The use of a strong form at the cæsura suggests that phrasing II is used here even in clause-internal position.

We can draw only limited conclusions from this data. Certainly, there is positive evidence that all Y-NOUS-LES pronouns could be stressed in the 12th century in clause-final position, where phrasing II was obligatory. Phrasing II is also attested in clause-internal position (at least for *nous* and *vous*). These

18. *c.*1180, ed. Roach (1959).
19. See Rydberg (1896–1907: 485–98) for further details on the development of this form.

findings corroborate the evidence from the MOI type of pronouns. However, no inferences can be drawn from clause-internal tokens which are not in metrically stressed positions, since either phrasing I or phrasing II could be used with no effect on the morphological form of the pronoun.

2.2.3 *Le*

The masculine singular direct object pronoun *le* is a special case because of its schwa vowel. This marks it out for two reasons. Firstly, it is not clear that the ModFr pronunciation of post-verbal *le* [lø] is medieval in origin.[20] Secondly, unstressed *le* is compatible with phrasing II without disrupting the regular PhPh-final stress rule. For example, *prends-le* could be pronounced /ˈprɛ̃nts.lə/: a single PhPh with stress on the final non-schwa syllable.

However, there is evidence not only that phrasing II was used for *le*, but also that *le* was stressed. While I am aware of no tokens of *le* at the rhyme in the 12th and 13th centuries, Melander (1928: 99–100) finds 20 tokens of *le* at the epic cæsura in his corpus of *chansons de geste*, e.g.:

(43) succurez **le** // a voz espiez trenchant
(*Chanson de Roland*,[21] l. 3378; Melander, 1928: 99)

(44) adobez **le** // a lei de chevalier
(*Couronnement de Louis*,[22] l. 1646; Melander, 1928: 99)

Here, as in examples (41–42) above, tokens of post-verbal pronouns at the cæsura are likely to be PhPh-final, and therefore show phrasing II. Moreover, post-tonic schwa syllables in polysyllabic words cannot occur at the epic cæsura: instead, the final non-schwa syllable of the word occurs at the cæsura, with the post-tonic schwa uncounted as at the end of the line. This evidence suggests that *le* was stressed at this time.

In both (43) and (44), phrasing II is used in contexts where phrasing I would also be possible. However, evidence from elision shows that phrasing II is not obligatory. In our corpus, there are two tokens from non-Picard texts where post-verbal *le* is elided before a vowel:

20. Foulet (1924) argues that *le* could be stressed, and assumes that the vowel has an [ø] realization, as in ModFr. However, since the /ø/ phoneme was not introduced into French until the 12th century (Pope, 1952: §§541–56), and there is good evidence for the alternation of [ə] and [e] in the pronunciation of the /ə/ phoneme in the 12th and 13th centuries (cf. chapter one, §3.1.2), I assume with Pope (1952: §275) that if *le* was stressed, a /e/ pronunciation seems more probable. Confusion with /le/ (plural) following the loss of final /s/ may explain its subsequent disappearance.

21. Ed. Bédier (1928).

22. Second third of the 12th century, ed. Langlois (1925).

(45) ot **le** O.li.ver, // sin ad mult grant i.rur (*ChRoland*, l. 1224)

(46) ge.tai **le** el Toi.(vre), // sel men.gie.rent poi.sson (*CharNimes*, l. 209)

For the MOI type of pronoun, tokens of this kind were only attested in Picard or in a restricted set of expressions. From these data, it would appear that phrasing I is more widespread with *le*. It is perhaps significant that the only token of post-verbal *le* before a vowel from the 14th and 15th centuries does not show elision:

(47) en.cu.se **le** a.per.te.ment (*PassPalat*, l. 398)

For the MOI type of pronoun, examples of pronominal elision before a vowel are restricted to a number of fixed expressions in the 14th and 15th centuries. It is possible that elision of *le* was also more restricted at this time; however, this hypothesis would need to be verified over a wider range of data.

We draw similar conclusions from the tokens of *le* as from the MOI and Y-NOUS-LES types of pronoun. Most importantly, there is clear evidence both for phrasing II and of a stressed realization of *le* when phrasing II is used. However, several tokens showing elision of *le* in contexts where phrasing I would also be possible suggest that a stressed realization of *le* is not obligatory. Moreover, such tokens are attested in extracts (*ChRoland*, *CharNimes*) in which strong forms of the MOI type of pronoun are obligatory in post-verbal position. It is therefore possible that unstressed realizations of *le* were more widespread than those of the MOI type.

2.3 Analysis

From the data, it is clear that phrasing II and stressed post-verbal pronouns of all types are present in MedFr from the 12th century. For Foulet (1924) and Ramsden (1963), the appearance of stressed forms in post-verbal position is evidence for the changed prosody of MedFr: Foulet (1924: 74) refers to an 'accent de groupe' at this time.

The 'accent de groupe' explanation for the emergence of stressed post-verbal forms is problematic. The data from *PassClerm* and *VieLeger* suggest that the strong post-verbal forms of the MOI type replace formerly unstressed forms. We thus suppose a historical alternation of the type **donez me* /dõ.ˈnets.mə/ vs. **donem* /ˈdõ.nəm/, with the schwa-final form following a consonant, and the consonantal form following a vowel. Foulet suggests that a 'lack of identity' of the pronoun in such combinations renders them impossible:

> Dans 'prenés le' l'enclitique s'appuiera nécessairement sur le verbe, mais ne pouvant perdre son identité dans la combinaison ainsi formée conservera sa voyelle et prendra l'accent du groupe.
>
> (Foulet, 1924: 72)

Yet such a lack of identity is precisely what is found for pre-verbal *le*, which remains consonantal before a vowel to this day. Moreover, even if the vowel of *le* is retained, this does not necessarily mean it will be stressed. Indeed, a regular group-stress rule is predicted to realize Foulet's /prẽnts.lə/ sequence with stress on /prẽnts/, rather than on the final /lə/ syllable. Equally, a group-stress rule could apply to /dõ.nets.mə/, and assign stress to /nets/, or to /dõ.nəm/, and assign stress to /dõ/. In short, the emergence of group stress cannot explain the generalization of the strong forms in the MOI type or the development of stressed *le*.[23]

The only rule that can explain the generalization of such forms is one which requires PhPhs to end in a PWd. Selkirk (1996) identifies exactly such a rule in modern English. She points out that when function words are PhPh-final in English, they must have a stressed realization:

(48) [What did you look at]$_{\text{PhPh}}$ [yesterday]$_{\text{PhPh}}$ ([æt], *[ət])

(Selkirk, 1996: 200)

Moreover, data from McCarthy (1991, 1993) show that the insertion of intrusive /r/ in English occurs only in two contexts: after a lexical word, and after a phrase-final function word (Selkirk, 1996: 200). This is further evidence that phrase-final function words are PWds. Selkirk thus proposes that there is a requirement in modern English for the PhPh to end with a PWd.

In §1.1.3, I argued that function words in MedFr are free clitics in Selkirk's (1996) model, and are not incorporated into the preceding or following PWd. Selkirk's analysis of modern English is thus directly applicably to MedFr. The **donez me* and **done m* forms are ruled out since the PhPh-final element is not a PWd:[24]

(49) *[(donez)$_{\text{PWd}}$ me]$_{\text{PhPh}}$

(50) *[(done)$_{\text{PWd}}$ m]$_{\text{PhPh}}$

23. It would, however, be sufficient to explain the emergence of stressed forms in the Y-NOUS-LES type, in which neither consonantal nor schwa-final forms are attested.

24. One possible criticism of this analysis is that it does not account for the attested cases of consonantal post-verbal pronouns in *PassClerm* and *VieLeger*. The data from these texts requires more careful analysis of consonantal forms in all contexts. However, while I maintain that French function words were free clitics in the 12$^{\text{th}}$ and 13$^{\text{th}}$ centuries, it is certainly possible that they were not in the early 11$^{\text{th}}$ century. Consonantal pronouns could thus have been incorporated into the preceding PWd. This remains an area for future research.

In the case of the MOI forms, the requirement for a final PWd is satisfied by substituting the strong form:

(51) [(donez)$_{\text{PWd}}$ (moi)$_{\text{PWd}}$]$_{\text{PhPh}}$

In the case of the Y-NOUS-LES forms, the requirement is satisified by parsing the only available form (the weak form) as a full PWd:

(52) [(vas)$_{\text{PWd}}$ (i)$_{\text{PWd}}$]$_{\text{PhPh}}$

In the case of *le*, the schwa-final or weak form is ruled out in the same way as the *me* or *m* forms. Thus, the only available form is a stressed weak form of *le*:

(53) [(prends)$_{\text{PWd}}$ (le)$_{\text{PWd}}$]$_{\text{PhPh}}$

This correctly predicts a stressed realization of *le* even though a final unstressed schwa syllable does not contravene the regular stress pattern of the PhPh. [25]

It is important to emphasize that this analysis explains only the *development* of stressed forms in post-verbal position. For the MOI type of pronoun, the prosodic rule is reanalysed as a morphological rule: thus in ModFr, strong forms of the MOI type invariably appear in post-verbal position, whether PhPh-final or not (e.g. *aidez moi donc* de Kok, 1985: 293). It is difficult to assess when this reanalysis may have taken place. However, we speculate that it may be linked to the restriction of post-verbal pronouns to imperative constructions over the course of the 13$^{\text{th}}$ century (cf. de Kok, 1985: 308–13). Tokens of weak post-verbal pronouns are found only in specific constructions in the 14$^{\text{th}}$ century (e.g. *vés me cy*, *laissez m'ester*), and could be considered morphological relics.

Crucially, the analysis we have adopted of the spread of post-verbal pronouns is clearly applicable to a language with word stress, modern English. We may therefore reject the hypothesis that the development of stressed post-verbal pronouns is positive evidence for the development of group stress.

One feature of the data which remains for future study is the use of phrasing I as opposed to phrasing II. Data from the MOI type of pronouns suggest that phrasing II is generally selected over phrasing I in most dialects in the 12$^{\text{th}}$ century. One possible explanation is that the obligatory use of phrasing II where the pronoun is clause-final (i.e. in imperatives) extends by analogy to

25. Post (2000: 92, note 20) extends Selkirk's analysis to final function words in the ModFr PhPh, in keeping with the assumption that primary stress in ModFr must be realized on PWds. In the framework of the present study, I argued in chapter one (§1) that ModFr primary stress is essentially a group-final stress rather than a word-final stress, and so there seems to be no reason for the function word to be parsed as a PWd. Certainly, unlike in the MedFr examples discussed above, such a rule has no effect on morphological form in ModFr.

cases in which phrasing I is also possible. Certainly, examples (14–18) from *PassClerm*, *VieLeger* and *Alexis* suggest that phrasing I was more widespread in 11th-century texts. Alternatively, it may represent part of a more general tendency for a monosyllable following the lexical head of a PhPh to undergo PhPh-restructuring. A full analysis of the causes of variation in phonological phrasing in MedFr adds little to our understanding of the emergence of group stress, and we will not pursue the issue here.

2.4 Conclusions

The data presented above suggest that strong forms in post-verbal position generalize in all except Picard texts before the 12th century. Moreover, exceptions to the rule are only attested in cases where the pronoun may not be PhPh-final. I have argued that stressed post-verbal forms emerge as a consequence of a rule requiring that the final element in a PhPh is a PWd. Such a rule is not the result of a group-stress system, as it is also attested in word-stress languages such as modern English.

However, the effects of this rule may have been vital in producing the conditions for the reanalysis of the MedFr stress system. The combination of regular final word stress and the requirement that every PhPh end in a stressable word has a crucial effect: every PhPh, regardless of its lexical or functional content, ends in a stressed syllable (or post-tonic schwa). While I have argued based on verse evidence that MedFr retained word stress in the 12th century, it is clear that this was a word-stress system which had also developed a regular PhPh-final stress. Coupled with the findings from section 1 of the present chapter, this reinforces the argument that the stress system of the 12th century showed all the features necessary to favour its reanalysis as a group-stress system.

3 The clause-initial constituent

In chapter one (§3.2.1), we considered the long-standing hypothesis that prosodic factors played a role in the verb-second system of French. These approaches were shown to be extremely problematic. However, in this section we will examine the rather different hypothesis that the emergence of group stress triggered syntactic changes which gave rise to a linguistic system in which the evidence for a verb-second constraint was less robust. Such a position is defended by Kroch (1989), and accepted with a number of reservations by Vance

(1997: 341–42), an analysis discussed in more detail in section 3.2.1. Such a finding would show that changes in prosody can have wider implications for the syntax of a language.

The analysis will focus on the prosodic, syntactic and pragmatic properties of short clause-initial constituents. In order to analyse syntactic changes, I will refer to generative analyses of MedFr word order, in particular the comprehensive studies of Adams (1987b) and Vance (1997), presented in §3.1. In chapter three (§1.2), we observed that the framework of Nespor and Vogel (1986) was difficult to apply to the clausal domain in MedFr, and in particular that the phonological phrasing of adverbial constituents was unclear. Since the phonological phrasing of initial constituents is vital to the present analysis, in section 3.2 evidence from versification will be used to clarify the predictions of the prosodic phonology model for MedFr.

Following this outline of the syntactic and prosodic assumptions of the analysis, I discuss the development of short initial constituents. As the subject pronoun and *si* have been extensively studied, I consider these only briefly, and focus instead on the adverbs *moult*, *bien* and *or*, examining whether there is evidence that the clausal position and pragmatics of these initial adverbs is affected by the emergence of group stress.

3.1 The clause structure of MedFr

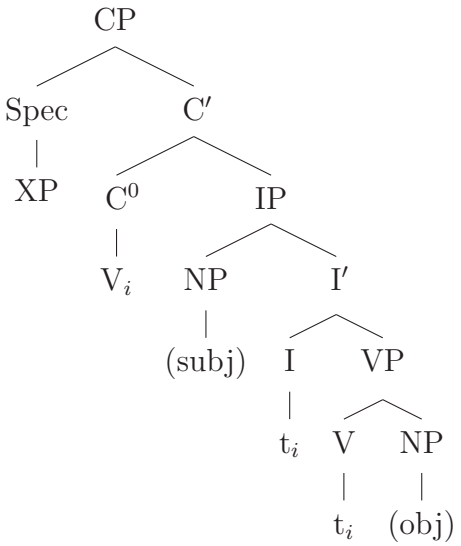
The present analysis assumes a broadly GB model of MedFr clause structure such as that first proposed by Adams (1987b) and best developed by Vance (1997). The model does not represent the 'state of the art' in generative syntax (cf. for MedFr Labelle and Hirschbühler, 2005; Mathieu, 2006; Labelle, 2007). However, for what is essentially a study of the interaction of syntax and prosody, the precise syntactic features and mechanisms used to derive word order are of less interest than the clausal architecture of the surface form. In this respect, there is less difference between the GB and minimalist analyses. A secondary advantage is that the three most comprehensive treatments of MedFr syntax (Adams, 1987b; Roberts, 1993; Vance, 1997) all assume a GB framework. Of these, Adams' model is the least complex,[26] and we will follow its basic structure except where it is inadequate to model the data relevant to our analysis.

26. Both Roberts (1993) and Vance (1997) assume a split-IP structure. Since I will be focusing exclusively on the main clause initial constituent, which I will assume to be in the C-domain, the articulation of IP is of little interest.

3.1.1 The Adams (1987b) model

The model of MedFr clause structure proposed by Adams (1987b) is as follows: [27]

(54)



(Adams, 1987b: 57)

Following Adams' analysis, the verb moves first into the IP (inflection phrase) to receive tense marking, and in main clauses then moves to the head of CP (complementizer phrase). A further XP (phrasal constituent) is then moved to Spec,CP (Adams, 1987b: 57–58). Although not shown on the tree diagram, this is assumed to have moved up from the VP (for direct objects, past participles, some adverbs) or the IP (subject, some adverbs). Clitics (negation, object pronouns) are incorporated into the finite verb at I^0, and then moved together with the verb to C^0. In subordinate clauses, the presence of a complementizer in C^0 prevents the verb raising beyond the I-domain, and thus verb-second effects are not predicted. [28]

This model makes important claims about the structure of main clauses in MedFr which can be used as diagnostics for syntactic structure. Firstly, one and only one constituent may precede the finite verb. The constituent has 'moved' to Spec,CP from within the clause. We saw in chapter one (§3.2.1) that a wide

27. The origins of this approach lie in work by Thiersch (1978) and den Besten (1983) on modern German.

28. This prediction is problematic in MedFr, since non-subject constituents are frequently attested before the finite verb in subordinate clauses, leading some authors to claim that V2 in both main and subordinate clauses must take place in the I-domain (e.g. Lemieux and Dupuis, 1995). I will not be considering data from subordinate clauses in the present study; however, Vance (1997: 148–66) highlights a number of asymmetries between V2 in main and subordinate clauses (in particular, the restriction of V2 in subordinates to the complements of 'bridge verbs' and a subset of complementizers) which are not compatible with an I-domain account of MedFr V2.

variety of constituents may occupy the pre-verbal position. However, none are restricted to this position, and may occur elsewhere in the clause. Secondly, where it is not pre-verbal (i.e. in Spec,CP), the subject, if expressed, is directly post-verbal.

3.1.2 The pragmatic function of the initial constituent

As discussed in chapter one (§3.2.1), Adams (1987b, 1989) argues that the constituent in Spec,CP is fronted to fulfill a prosodic requirement and has 'no necessary marked or emphatic interpretation' (1989: 4), unless it is the pre-verbal object. We rejected this explanation of V2 word order.

The most common explanation for constituent fronting is precisely that rejected by Adams (1987b, 1989): the initial constituent has a marked discourse function. Vance (1997: 28) claims that Spec,CP is a topic position, while in later generative work, Labelle and Hirschbühler (2005: 61) assume a broader discourse function, claiming that the initial constituent may be either topicalized or focalized. The idea that V2 is a 'topic-initial' system of word order is also widespread in studies of French syntax outside the generative framework.[29] Marchello-Nizia (1995: 86–89) and Vance (1997: 233–44) suggest that the initial constituent in the 13th century may have a 'linking' role, referring back anaphorically to elements present in preceding discourse. Marchello-Nizia (1995: 89–92) also identifies a number of cases in which an initial direct object may be topicalized without referring to previous discourse (a cataphoric use), and a separate case in which it is fronted for emphasis. We conclude that there are clear cases in which the initial constituent is fronted for pragmatic reasons.

Yet as a general explanation for constituent fronting in a V2 system, this too is insufficient. In 13th-century prose, Vance (1997: 40–41) observes that initial subjects are rarely linked to the preceding discourse, and does not treat them as pragmatically marked.[30] However, in other analyses, it commonly assumed that subjects are the *most common* topics: this is how Vennemann (1974), Harris (1978) and Marchello-Nizia (1995: 100) explain the development of SVO word order from a V2 system, which they characterize as TVX (Topic–verb–other material). More problematic still, this lack of pragmatic marking is not restricted to subjects. Marchello-Nizia (1995: 97) argues that in the *Chanson de Roland*, pre-verbal objects are not pragmatically marked: instead, this word order is regular in null subject clauses. This word order cannot be

29. E.g. Harris (1978), Buridant (1987, 1992), Marchello-Nizia (1995, 2006).

30. Vance (1997) assumes that SV clauses are IPs, rather than CPs in the 13th century. However, I do not adopt this position, as the evidence is ambiguous, see §3.1.3 below.

predicted by a generative V2 model in which constituents are only fronted for pragmatic reasons. I suggest that while pragmatic effects are common, they do not determine the syntactic requirement for an initial constituent.

3.1.3 Verb-third exceptions

Verb-third word orders in main clauses are in principle prohibited by the model of a V2 language suggested above, since there is only one pre-verbal position (Spec,CP). However, they are clearly attested in MedFr. I will divide the discussion of verb-third examples into two types, each with different analyses. Type I V3 is the only type attested in the 12th and 13th centuries, and is compatible with the V2 model. Type II V3, on the other hand, is only attested from the 14th century, and I will argue that it represents a new, non-V2 word order. [31]

In the 12th and 13th centuries, verb-third (Type I) in main clauses is rare but attested (Adams, 1987b: 190–92; Roberts, 1993: 95; Vance, 1997: 59–66; Lemieux and Dupuis, 1995: 100). Vance's 13th-century prose data contains tokens of a word order that she characterizes as CSV: [32]

(55) et [neporec] [il] le diroit volontiers
(*Queste del saint Graal*, [33] p. 66, l. 1; Vance, 1997: 62)

Vance claims that adverbials such as *neporec* and *certes* never trigger verb–subject inversion. Indeed, they can also precede CVS clauses:

(56) Et [neporquant] [de la force de lor lances] arestent il son cheval en plain cors
(*ibid.*, p. 48, l. 3; Vance, 1997: 66)

In the GB framework, such examples are reconciled with the V2 model by assuming that the adverb is dislocated and falls outside the clause (Adams, 1987b: 194; Vance, 1997: 61). A more sophisticated treatment is provided in recent generative work, based on the idea that the CP actually consists of a series of projections with different pragmatic functions (Rizzi, 1997). The following map of the C-domain is proposed by Benincà and Poletto (2004), and presented for medieval Romance by Benincà (2006: 61):

31. A similar division is made by Adams (1987b: 191–96).
32. I use the traditional philological / typological abbreviations: O = object, S = subject, V = verb, X = intervening material, C = complement. Additionally, following Vance (1997): Sp = pronominal subject, Sn = noun phrase subject.
33. *c.*1220, ed. Pauphilet (1923).

(57) [Force C^0
{$_{\text{Frame}}$[ScSett][HT] C^0 }
{$_{\text{Topic}}$[LD][List] C^0}
{$_{\text{Focus}}$[EmpFocus][UnmFocus] C^0} [Fin C^0]]

(adapted from Benincà, 2006: 76)

The braces denote fields (groups of similar projections), while the brackets denote individual projections. Force and Fin are complementizer positions. The highest field (Frame) contains scene setting adverbials (ScSett) and hanging topics (HT). The middle field (Topic) contains left-dislocated topics (LD) and listed elements (List).[34] The lowest field (Focus) contains elements with emphatic or unmarked focus.

The analysis makes interesting predictions relating to constituents in the Topic and Frame fields. Pragmatically, constituents in these fields cannot be focalized, they must be dislocated topic constituents or 'scene setting' adverbials. Moreover, elements in the Topic and Frame fields must be merged in this position, not moved from within the clause, while the same restriction does not apply to the Focus field (Benincà, 2006: 77). Therefore, we predict that fundamentally clause-internal elements cannot move to these positions. Thus, for example, if a noun phrase representing the direct object is located in the Topic or Frame fields, it is predicted to be doubled by a pronoun, fulfilling the direct object function within the clause (i.e. clitic left dislocation, cf. Benincà, 2006).

As with previous analyses, I assume that Type I V3 is fundamentally a V2 order: the verb is in C^0 and the immediately preverbal constituent in Spec,CP.[35] The first element must be in the Frame or Topic field, and must show the pragmatic and syntactic features of such elements. For a phrasal adverb such as *neporec*, this is the clearly the best analysis: it is neither focalized nor clause internal. Moreover, Type I V3 is not fundamentally a CSV order but an XCV order: any V2 initial constituent may occupy the immediately pre-verbal position.

In the 14$^{\text{th}}$ and 15$^{\text{th}}$ centuries, however, a verb-third word order begins to be attested which shows entirely different properties (Type II):

(58) Et [**ce conseil**] [nous] vous donnons

(Jean Froissart, *Chroniques*; Adams, 1987b: 195)

34. For the distinction between left-dislocated and hanging topics, see Benincà (2001: 43–44). Essentially, both types of constituent contain clause-internal pronominal copies, but hanging topics, unlike LD topics, may not be preceded by a preposition.

35. I assume that the 'CP' of GB analyses is the lowest projection in the C-domain to contain lexical material. Whether this is consistently Spec,FocP, as Labelle (2007) suggests, or whether its position varies within the split-CP depending on the pragmatic value of the initial constituent (Benincà, 2006) is open to debate.

The direct object *ce conseil* is clause-internal, and thus cannot occur in the Topic or Frame fields. Out of context, the pragmatic reading of this example is uncertain; however, a 14th-century example from the prose subcorpus shows that an emphatic reading is possible:

(59) Et se Famius en avoit grant angoisse et douleur, encore en avoit sa femme sept tans plus, ne elle n'avoit joye ne repos par nuit ne par jour, ains estoit adés en oroisons, en depriant Dieu que elle en peüst un avoir, **ne [nulle autre priere] [elle] ne faisoit a Nostre Seigneur.** (*Berinus*, §9)

Here, the fronting of *nulle autre priere* emphasizes the fact that the only wish of Famius and his wife is to have a child. Taking the pragmatic and syntactic evidence together, here the direct object is fronted to the Focus field, just as in V2, but is not immediately followed by the verb.

Adams (1987b: 196) argues that this structure is only possible with pronominal subjects, and moreover that it is still a V2 structure. She claims that subject pronouns in the 14th and 15th centuries are 'optionally clitic' on the finite verb. Yet the notion of 'optional cliticization' is virtually unfalsifiable, since in verb-third structures, the subject pronoun is analysed as a clitic in C^0, and in verb-second structures as a full pronoun in Spec,CP. In a detailed refutation of Adams' analysis, Vance (1997: 282–88) argues that while pronominal subjects are more common here, nominal subjects are also attested, for example:

(60) [le petit Saintré] [les yeulz de Madame] ne cessoient de regarder, tant danssoit et chantoit bien

(*Jehan de Saintré*;[36] Vance, 1997: 283)

Tokens of this kind are certainly rare: in the syntactically tagged subcorpora of my own corpus, none were attested in 14th- and 15th-century texts, against seven tokens of main clause CSpV. However, Vance speculates that the rarity of nominal subjects in this position may be attributed to a rhythmic restriction, limiting the number of stressed pre-verbal XPs to one (Vance, 1997: 284).[37]

We will adopt Vance's view that subject pronouns are not syntactic clitics in the 14th and 15th centuries, and thus examples such as (58) and (60) represent a new type of word order, one in which the verb and the subject remain in IP, leaving Spec,CP free for the first element in the verb-third structure:[38]

36. Ed. Knudson and Misrahi (1965).

37. A similar view is developed at length for the history of English by Speyer (2010). The rhythmic factors in question, however, operate at a higher level on the prosodic hierarchy than the change we are investigating.

38. Vance (1997: 270) suggests that the initial topic is adjoined to IP in the 15th century,

(61) [$_{CP}$ le petit Saintre [$_{IP}$ les yeulz de Madame [$_{I'}$ ne+cessoient ...]]]

The fundamental characteristic of a Type II V3 structure is that the verb remains in the I-domain. This is not therefore a V2 word order. Derived from this are two further properties: the initial element may be a fronted and emphatically focused clause-internal constituent, and only the subject may intervene between the first element and the verb. Vance (1997: 341–47) argues that it is the increase in CSV word orders which undermines the V2 system of MedFr, leading to its eventual disappearance. However, the reason for the increase in CSV structures of this type is unclear. We will return to this issue in section 3.3.1.

The distinction between Type I and Type II V3 is not always clear-cut. Vance (1997: 63–66) notes an 'unstable' class of initial elements in 13th-century prose, which may occur with either CVS or CSV word order. Primary among these in Vance's corpus are *onques* and *ja*. In her analysis, where CSV word order is attested after these adverbials, the verb remains in the I-domain. In our typology, this would therefore be a Type II V3 order. However, the analysis of these constructions is problematic. There seems no pragmatic reason to suppose that *onques* or *ja* must be fronted from clause-internal position rather than base generated in the Topic or Frame fields. Moreover, there are tokens in my corpus from 13th-century verse of *onques* and *ja* in XCV rather than CSV structures:

(62) [ja] [de mon gent cors] n'arés part (*ComtePoit*, l. 220)

(63) car [onques] [en cel lieu] ne vint
k'il trovast qui bien li desist (*ChvBarisel*, l. 650–51)

In my view, these examples suggest that Vance's tokens in 13th-century prose may still be of Type I rather than Type II V3. I conclude that the emergence of the non-V2 Type II V3 order cannot convincingly be dated to the first half of the 13th century. Moreover, the examples demonstrate that short adverbs may occupy a variety of clausal positions (post-verbal; Spec,CP; Spec,FrameP). Because of the focus here on the role of prosody in word order change, in section 3.3, I will focus on adverbs with less variable syntax. The precise syntax of many short adverbs in MedFr remains a topic for future research.

rather than in Spec,CP. In favour of this analysis, she notes the frequent possibility of CSV in embedded clauses. However, it is not necessarily that case that the analysis of embedded and main clauses must be identical (cf. the approach of Labelle (2007) for MedFr of the early 12th century), and I retain a CP analysis.

3.2 V2 and prosodic change

3.2.1 Kroch (1989)

Kroch (1989) gives the most satisfactory account proposing a link between prosodic change and the loss of V2 in MedFr. Kroch argues that the loss of verb-second effects in the 14th and 15th centuries can be attributed to a weakening of phrase-initial stress. By 'phrase', Kroch is referring to the 'major intonation phrase' (1989: 214), a level higher on the prosodic hierarchy than the PhPh. Kroch assumes that topics must be stressed. Thus, clause-initial topics become impossible following the loss of phrase-initial stress in French. Instead, topics are left-dislocated, moving outside the clause and forming a separate intonation phrase (Kroch, 1989: 214). Kroch argues that left-dislocated constituents require a clitic copy within the clause, as illustrated in the following ModFr example: [39]

(64) (a) *Je déteste les courgettes, mais les haricots$_i$, j'adore e$_i$.
(b) Je déteste les courgettes, mais les haricots$_i$, je les$_i$ adore.

(Kroch, 1989: 213)

Since left-dislocated topics cannot fulfill the V2 requirement, another constituent, usually the subject pronoun, must be moved to Spec,CP. This leads to an increase in TopSV word orders. However, as TopSV constructions are equally compatible with an SVO grammar, over time the evidence for a verb-second grammar is weakened, and V2 effects disappear. To support this claim, Kroch demonstrates that the decline in verb–subject inversion and the decline of null subjects, both features of the V2 grammar of MedFr, proceed at the same rate as a rise in clitic left dislocation (cf. Priestley, 1955).

Vance (1997: 341–42) argues that a rhythmic account of this type could potentially explain the increase in CSV constructions in MedFr. On this point, I agree with Vance. However, I am less convinced by Vance's reliance on Kroch's account, as the key changes occur too late to explain the rise in CSV. Kroch's account focuses on the *loss* of verb–subject inversion and null subjects, processes which date from the end of the 15th century and into the 16th century (cf. Fontaine, 1985). Priestley's (1955) data show no major increase in the rate of clitic left-dislocation between 1200 and 1450, but the rate more than doubles between 1450 and 1550 (Kroch, 1989: 215). However, Vance (1997: 277) argues

39. In modern spoken French, however, (64a) is a perfectly grammatical topicalization, cf. Blanche-Benveniste (1996). Kroch (1989) does not discuss this problem, although if his argument is correct, the grammaticality of (64a) must be regarded as a modern innovation following earlier generalization of the reprise construction (64b).

that CSV word order is already attested in Joinville's *Vie de saint Louis* from the early 14th century, and is 'abundant' in Froissart's *Chroniques* from the late 14th century. In short, Kroch's rhythmic account cannot explain the rise in CSV constructions in the way that Vance suggests.

Vance (1997: 342) herself points out a second difficulty for Kroch's account. Kroch hypothesizes that topics move out of the clause because they require stress. Yet it is not just nominal topics that appear in CSV configurations, but also short adverbs such as *lors* and *or* (at least in Vance's data) from the 14th century. In Vance's view, it is implausible that these short elements should require stress in the same way as a topicalized noun phrase. The rhythmic account thus wrongly predicts a 'split' in the class of initial constituents:

> A plausible development might see a stable split between 'heavy' topics, which fail to trigger V2 (opting for extraposition or 'reprise') and 'light' topics, such as adverbials, which would continue to be acceptable in the topic position [...] Spec,CP.
>
> (Vance, 1997: 347)

3.2.2 A new prosodic constraint

To resolve some of the problems for the rhythmic account, I will assume a slightly different view of the relationship between stress, syntax and pragmatics. Kroch (1989) and Vance (1997) assume that initial constituents are stressed because they are topics, and topicalization requires stress. Yet we saw in §3.1.2 that not all initial constituents are pragmatically marked, a point we will develop further in section 3.4. As Adams (1987b, 1989) points out, the initial constituent in modern V2 languages is always stressed, whether topicalized or not. Like Adams, I propose that stress on the initial constituent is not a consequence of topicalization. Instead, I suggest that it is an inherent property of any constituent fronted to Spec,CP:

> **Stress condition**
> XP constituents moved to CP are realized as stressed.

This is crucially different to Adams' suggestion, which is that a stress requirement triggers movement. In my approach, it is the other way around: constituent fronting triggers stress.[40] Crucially, pragmatic function and stress

40. This leaves the constituent fronting which characterizes V2 word order unexplained, but it is perhaps unlikely that there exists a deep pragmatic, prosodic or syntactic 'reason' for basic word orders of this kind. Moreover, cases of obligatory but 'unexplained' movement are not uncommon in generative approaches. In particular, movement may be triggered by

are independent, thus I allow both for unstressed elements with pragmatic function and for stressed elements with no pragmatic function. We will see both of these attested in MedFr. Moreover, we predict that any loss of stress on the initial constituent will initiate linguistic change, as learners adopt an analysis in which the constituent is not fronted to Spec,CP.

3.2.3 The phonological phrasing of the initial constituent

For Kroch (1989), prosodic change in MedFr causes initial elements in the intonation phrase, and thus all clause-initial elements, to become unstressed. Indeed, the effects of this change are most clearly observed if clause-initial noun phrase topics are considered. In the present thesis, however, we have found little evidence to suggest that the emergence of group stress causes stress to be lost on noun phrases. The head noun of a noun phrase is frequently predicted to be PhPh-final, and thus remains a bearer of primary stress into ModFr.

The emergence of group stress is predicted to affect those initial constituents which do not form their own PhPh. In chapter three (§1.2), we saw that the subject pronoun and the adverb *si* are incorporated into the PhPh containing the finite verb. One of the key features of a group-stress system highlighted in chapter one (§1) and again in section 1 of the present chapter is the lack of lexically-associated stress realized on non-final elements in the PhPh. We conclude that the emergence of group stress caused the loss of the word stress on all clause-initial constituents which were incorporated into the PhPh containing the finite verb. I will henceforth refer to these as 'light' initial constituents.

In chapter three (§1.2), we noted that the phonological phrasing of adverbs was difficult to reconstruct. In this section, however, I will assume that monosyllabic initial adverbs were light initial constituents.

Evidence from versification can be used to support this claim. I have assumed that the cæsura and the rhyme correspond to a PhPh-boundary (cf. chapter one, §4). Therefore, if constituents end regularly at the cæsura or the rhyme, this suggests that they are PhPh-final. Clause-initial noun phrases, for example, are attested at the cæsura, much as we would predict:

(65) **Eufemien** // volt saveir quet espelt (*Alexis*, l. 350)

(66) et **Damedix** // en soit hui grasiiés (*CharNimes*, l. 401)

an EPP feature. Holmberg interprets the EPP as containing a [P] feature (short for 'phonological'): 'an uninterpretable feature checked by a phonologically visible category moved to or merged in [the specifier position]' (2000: 456). Applying this to the V2 case, a constituent moves to Spec,CP because Spec,CP must be realized phonologically.

(67) **chrestientét** // aidez a sustenir! (*ChRoland*, l. 1129)

Clause-initial subject pronouns and *si*, on the other hand, are never attested in this position. In part, this may be simply a matter of length. Clause-initial noun phrases may be polysyllabic, and hence capable of filling the whole of the first hemistich. However, a clause-initial monosyllable must be preceded by clause-external material (such as a vocative) in order to occur at the cæsura.

This problem may be partly resolved by including tokens of pre-verbal subject pronouns from subordinate clauses. Four tokens of subject pronouns in fourth position in a ten-syllable line are attested:

(68) sai.ciés ke **il** // li fau.ra a gri.gnor (*CononBeth*, IV, l. 12)

(69) de chou ke **vous** // m'a.vés fait en.du.rer (*AdamHale*, IV, l. 30)

(70) un che.va.lier // qui pour e.lle a.ffo.lloit,
avant qu'**e.lle** // l'a.mast tant se dou.loit (*3Jugemens*, ll. 41–42)

(71) et pour ce qu'**e.lle** // i.roit sans or.den.an.(ce) (*OrlogeAmor*, l. 206)

There are two possible explanations for these examples. Either the subject pronoun exceptionally occurs at a PhPh edge, or the poet does not consistently mark the fourth position with a PhPh edge. In the case of (71), Billy's (1999) study of the versification of *OrlogeAmor* clearly demonstrates that the position of the PhPh edge in the line is not regular in this text. In particular 6+4 divisions are also attested. This seems to be a much more plausible analysis of this example. For (69) and (70), while there is always a word boundary after the fourth syllable in the line in these texts, it is not always a PhPh edge. In both texts, there is one token of an object pronoun in fourth position:

(72) mer.chi, da.me, // la cui biau.tés sour.vaint
mon cuer ki **vous** // a fait lo.ial ho.mma.(ge) (*AdamHale*, IV, ll. 33–34)

(73) et vous pro.met // que ja.mais au.tre a.tten.(te)
n'a.ray qu'a **vous** // ser.vir, car dou.lce ren.(te)
m'en pay.e.ra (*3Jugemens*, ll. 170–72)

Additionally, in *3Jugemens* there are four tokens of *qui* heading a subordinate clause in fourth position: again, the complementizer is very unlikely to occur at a PhPh edge. In (73), as in (70), there is a PhPh edge after the sixth syllable. This suggests that the 6+4 variant of the ten-syllable line described by Billy (1999) for *OrlogeAmor* may have been used by other authors in the late 14th and early 15th century. We conclude that the use of the subject pronoun in fourth position in (69–71) results from a different treatment of the cæsura rather than that the subject pronoun could be PhPh-final.

Example (68) may indicate a PhPh-final use of the subject pronoun. It is possible that the subject pronoun in (68) may be read with contrastive focus:

(74) por li m'en vois sospirant en Surie,
car je ne doi faillir mon Creator.
Ki li faura a cest besoig d'aïe,
saiciés ke il li faura a grignor (*CononBeth*, IV, ll. 9–12)

I read the couplet in 11–12 here as 'Whosoever fails him [God] in his need of help, know that he [God] will fail him still more.'[41] This is the reading present in manuscripts MRTa of Wallensköld's (1921) edition. However, a variant reading in manuscripts KNPVX avoids the construction, and clearly contains a PhPh boundary in fourth position:

(75) sai.chiés **de voir** // q'il fau.dra a gri.gnor

To summarize, subject pronouns may only be used at the cæsura when emphasized, or in texts where a PhPh edge is not required in this position. This confirms the view taken in chapter three (§1.2) that they are not PhPh-final.

Like subject pronouns, monosyllabic adverbs are also avoided at the rhyme and at the cæsura. In fact, single monosyllabic pre-verbal adverbs are unattested in these positions in the corpus. Such a restriction does not apply to all adverbs. Indeed, in the early-8 and base subcorpora alone, there are eight tokens of pre-verbal adverbs at the end of the line, for example:

(76) "seingnors, dist ele, **volentiers**
ert retenuz li chevaliers" (*Thebes*, ll. 291–92)

(77) et li chevax **eneslepas**
saut en l'eve et del chanp se soivre (*Charrete*, ll. 754–55)

In addition to *volontiers* and *eneslepas*, the other six adverbs are *proprement*, *communement*, *premierement*, *longuement*, *contremont* and *derechief*. All are polysyllabic, and all are lexically-derived. Disyllabic adverbs are rarer but attested at the rhyme (three tokens) and at the lyric cæsura (one token), e.g.:

(78) or nos mostre et dit salutes
las juenes chevaliers entrutes
et aslites, car quant **jaidis**
estoient puissant et hardis

41. However, the sense here is not entirely clear: *li* in line 9 refers to the lady the knight is leaving behind, but in ll. 11–12 the referents of *il* and *li* are less clear; as indeed is the reading of *faura*, which in line 12 could derive from either *falloir* or *faillir*.

Constituent	before 1250 (%)	after 1250 (%)
Sp	9.2	19.1
Ce	4.4	5.9
M/syll Adv	24.9	21.7
Overall	38.5	46.6
Sample	12 texts, 1955 tokens	15 texts, 1740 tokens

Table 5: Proportion of light initial constituents in surface verb-second main clauses, early-8, base and prose subcorpora

de combatre et de bataillier
et de ferir et de maillier,
il aprenoient as aberges
chevalerie per usaiges
de laborer, de travaillier,
per tresnutier et per vaillier. (*AbrCheval*, ll. 481–90)

My reading of the full sentence does not suggest a particularly emphatic or contrastive reading of *jaidis* here.

To summarize, while I am cautious of reading too much into negative evidence, it is suggestive that monosyllabic adverbs are not attested at the rhyme or the cæsura, while longer adverbs are. I suggest that it is very likely that short adverbs were not PhPh-final, and thus like subject pronouns were 'light' initial constituents.

3.3 Change and light initial constituents

In this section, we will see that the Stress Constraint proposed above makes correct predictions about the development of light initial constituents. I will briefly review the well-documented development of the subject pronoun, before proceeding to look at light initial adverbs.

Table 5 presents a breakdown of the overall proportion and individual types of light initial constituent in MedFr. The only significant trend in the data is the increase in the use of the subject pronoun. [42] As a brief overview, the table demonstrates that light initial elements are found in approximately 40% of

42. Regression of proportion of verb-second main clauses with initial subject pronoun on date of composition: $r = 0.443$, $t = 2.47$, $p > |t| = 0.021$.

all surface verb-second clauses in MedFr, and that the most common initial constituents are short adverbs. These elements all have an etymological word stress, which I assume was sufficient to conform to the Stress Constraint. This stress would have been lost with the emergence of group stress, and as a consequence, around 40% of initial constituents would no longer have conformed to the Stress Constraint. This is predict to have had a major effect on the syntax of the language.

3.3.1 Reanalysis of subject pronouns

Subject pronouns in the 12th century are assumed to be V2 initial constituents in Spec,CP (Adams, 1987b; Roberts, 1993; contra Vance, 1997). However, we saw above that the emergence of Type II V3 word order in the 14th century shows that by this point SV word order has been reanalysed as non-V2, with the verb and subject in the I-domain. In the prose subcorpus, this order is first attested in *Berinus*:[43]

(79) ne [nulle autre priere] [elle] ne faisoit a Nostre Seigneur.
(*Berinus*, §8)

(80) ne [joye ne bien] [ilz] ne pouoient avoir (*Berinus*, §9)

We saw in §3.1.3 that the object here was focused, suggesting that it is fronted to the Focus field of the CP with the subject and verb remaining in IP:

(81) ne [$_{\text{FocP}}$ nulle autre priere [$_{\text{IP}}$ elle [$_{\text{I}'}$ ne+faisoit ...]]]

In the present account, as in that of Roberts (1993), the existence of Type II V3 word orders in the 14th century is assumed to be a consequence of the reanalysis of CP-SVO as IP-SVO. However, in my view, the main trigger for this reanalysis is prosodic. According to the Stress Constraint, subject pronouns in Spec,CP must be stressed. The same restriction does not apply to Spec,IP. The loss of stress on subject pronouns in the 13th century due to the emergence of group stress favours the reanalysis of subject pronouns as

43. In the base subcorpus, OSV word order is common in the earliest texts, and remains, although rarely, in texts from the 12th century (Marchello-Nizia, 1995: 97). Two tokens from *PassJongl* are present in our corpus:

(i) [glaive] [chascun de vous] prendroiz (*PassJongl*, l. 430)

(ii) [de la paor que Dieus avoit]
goutes de sanc [sa char] suoit (*PassJongl*, ll. 491–92)

Both tokens are verb-final (indeed, 43 is verb-fourth). Given the potential for archaic features in the text (cf. chapter two, §1.4), I suggest that these represent a separate, archaic type of verb-final construction and not the earliest tokens of Type II V3. An analysis of these constructions must be left for future research.

Adverb	Tokens	Before 1250 (% total light Adv-initial clauses)	After 1250 (% total light Adv-initial clauses)
si*	286	29.5	37.6
or	77	7.6	10.6
mult	62	10.5	2.9
bien	61	8.8	4.7
puis	46	7.6	2.4
lors	41	4.3	5.3
tant	40	2.7	7.2
donc	38	5.5	2.9
ja	29	3.5	2.2
la	28	2.5	4.2
ains	27	2.5	4.0
plus	16	1.2	2.7
tout*	14	1.2	2.1
trop	14	0.4	3.2
ci	10	0.8	1.6

* Only adverbial tokens of *si* and *tout* are counted.

Table 6: Frequency of most common light adverbial initial constituent in main clause declaratives in the base and prose subcorpora

IP-internal, in order for the construction to conform to the Stress Constraint. Following this reanalysis, Spec,CP is left empty for other (stressed) arguments, giving rise to CSpV word order. CSnV word order, I assume, develops perhaps at a later date by analogy with CSpV.

3.3.2 Frequency of light initial adverbs

Table 6 shows every light initial adverb attested in the early-8, base or prose subcorpora with an overall frequency of at least ten tokens.[44] The tokens column gives the overall number of tokens of each adverb in all three subcorpora across the time period. The remaining two columns show the frequency of the adverb in initial position as a proportion of all the main clauses beginning with light initial adverbs. Thus, *si* is the initial constituent in 29.5% of clauses

44. Attested with fewer than ten tokens are *mieux* (8), *onques* (6), *mal*, *mar*, *tost*, *anc* (3), *peu*, *jus*, *hanc* (2), *fors*, *pres*, *pis*, *loin*, *sus* (1).

beginning with light initial adverbs from before 1250, but in 37.6% of such clauses from after 1250.

From table 6, it is clear that there are substantial differences in the frequency of individual adverbs in the pre- and post-1250 periods. Using a χ^2 test, significant differences are attested in the frequency of *si*, *tant* and *trop* (increase) and *moult*, *bien* and *puis* (decrease) in the two time periods.[45]

The most common adverb overall, *si*, is proportionally more common in the post-1250 period, while the second most common adverb (*or*) shows no significant change in frequency. However, *moult* and *bien*, both more common than *or* in the pre-1250 period, become significantly less common in later texts. In the following sections, I will examine the syntactic and pragmatic properties of these adverbs to investigate the reasons for their divergent developments. I will suggest that a reanalysis in keeping with the Stress Constraint following the emergence of group stress played an important role.

3.4 *Moult* and *bien*

In this section, I discuss the syntactic (§3.4.1) and pragmatic (§3.4.2) properties of *moult* and *bien* fronting in the 12th and 13th centuries, arguing that the adverbs are fronted for purely syntactic reasons to fulfill the V2 requirement for a pre-verbal constituent. The frequency of fronting declines in the 14th and 15th centuries, which I attribute to the need to conform to the Stress Constraint. Finally, we will see that remaining tokens of fronted *moult* and *bien* in the 14th and 15th centuries show very different syntactic and pragmatic properties (§3.4.3).

The historical development of *moult* is studied in detail by Marchello-Nizia (2000, 2006), along with the rise of its replacements *beaucoup* (as a quantifier) and *tres* (as an intensifier adverb). She demonstrates (2006: 140–41) that the decline of *moult* with respect to *tres* begins in the 14th century, but it is not until the late 15th century that *beaucoup* is attested in all texts, usually in competition with *moult* (2006: 154). Since the loss of *moult* largely post-dates our period of study, the explanation for the decline of *moult* as a pre-verbal constituent is to be sought elsewhere. Partly because Marchello-Nizia's study provides data to supplement the analysis, I will focus rather more on *moult* than *bien*, although I assume that the two adverbs had very similar properties.

45. *Si*: $\chi^2 = 4.21$, $p = 0.040$. *Tant*: $\chi^2 = 9.26$, $p = 0.002$. *Trop*: $\chi^2 = 10.08$, $p = 0.001$. *Moult*: $\chi^2 = 16.90$, $p = 0.000$. *Bien*: $\chi^2 = 4.95$, $p = 0.026$. *Puis*: $\chi^2 = 10.84$, $p = 0.001$. All with one degree of freedom.

3.4.1 Syntactic properties

Initial *moult* and *bien* show a number of clear syntactic properties. Firstly, there is evidence that they cannot occur higher in the clause than Spec,CP in the pre-1250 period. In tokens of Type I V3, the adverb is always in the immediately pre-verbal position: [46]

(82) cio fud lusos ut il intrat;
clerj' Ewrui illo trovat.
[Cil Ewruins] [**molt**] li vol miel (*VieLeger*, l. 101)

(83) [ciest omne tiel] [**mult**] aima deus (*VieLeger*, l. 207)

(84) de Deu loër ne se ublïent,
mais [sa merci] [**mult**] la crïent (*Brendan*, ll. 305–6)

In (82) and (83), the first element is a left-dislocated anaphoric topic, while in (84) it is a topicalized direct object: note the clitic copy of the dislocated object within the clause. Prosodically, the pre-verbal adverbs may be assumed to be within the PhPh of the verb: however, at this stage of the language, we predict that they would still have carried a word stress, thus conforming to the Stress Constraint. [47]

Secondly, *moult* and *bien* occur only very exceptionally occur with pronominal subjects. The data in Offord (1971) suggest that this holds true even in the 14th century: he finds only one token of a post-verbal subject pronoun after fronted *bien* compared to 322 with a null subject. In my own corpus, *moult* and *bien* are only attested with post-verbal nominal subjects. [48] Vance (1997: 66–126) presents a detailed analysis of the position of post-verbal subjects in the 13th century, arguing that nominal subjects are frequently lower in the clause than Spec,IP. Where the past participle, infinitive or direct object separate the subject from the verb, Vance argues that the subject remains in the VP. Where the subject is particularly long, Vance argues that it is clause-final. In the early-8, base and prose subcorpora, there are four tokens of *moult* where

46. This section focuses only on single adverb constituents. *Moult* and *bien* can both be separated from the finite verb if fronted as part of an XP constituent of any kind, such as *mult volontiers* or even *mult bien*. These are discounted since *moult* is not itself a constituent.

47. For *moult*, there is one element that can intervene between the adverb and the finite verb: the particle *par* (cf. Buridant, 2000: §439)

(i) iceste gent fole esbaie
mult par i firent grant folie (*GormIsem*, ll. 155–56)

I do not interpret *par* as a full constituent here, and thus (i) is not a counter-example to the generalization that *moult* must be immediately pre-verbal. However, as it is less clear that *moult par* is a light initial element, we will not consider such tokens further.

48. There is, however, one token of *moult par* with a subject pronoun.

the subject is not directly postverbal, and one of *bien*:

(85) mout demenerent [grant dolor]
[**cil de Thebes**] pour lor seingnor (*Thebes*, ll. 147–48)

(86) bien fust [a oés un roi] [metables]
[**li covertors qu'il ot sor lui**] (*Charrete*, ll. 510–11)

The fact that *moult* fronting is attested with nominal subjects but not with subject pronouns may be due to differences in the syntax of the two types of subject. However, while this generalization holds true in most texts, there are exceptions: tokens from *LaisMarie* (4), *RoseMeun* (1) and *Florimont* (1) show that CVSn order is possible when the nominal subject is high in the clause:

(87) **mut** est **Lanval** bien herbergez (*LaisMarie*, l. 140)

(88) **bien** le m'avoit **Reson** noté (*RoseMeun*, l. 4117)

(89) **bien** iert **ta fille** mariee
quant a ton signor iert donee (*Florimont*, ll. 1325–26)

In each case, the nominal subject precedes the past participle, indicating a position in the I-domain rather than in the VP.

Thirdly, where the subject pronoun is present, *moult* and *bien* appear in post-verbal position:

(90) vassal, vus me avez **mut** mesfait! (*LaisMarie*, l. 363)

(91) Sire cevaliers, je voi **bien** que vous m'avés abatu.
(*TristanPr*, p. 5, l. 26)

Finally, where *moult* and *bien* are both present in a clause, the corpus contains tokens of *moult*–V–*bien*, but not **bien*–V–*moult* (cf. also 87):

(92) bien semble houme de grant parage,
mout li plest **bien** en son corage. (*Thebes*, ll. 307–8)

The fronting of *moult* and *bien* thus seems to be very sensitive to syntactic constraints. In particular, there seems to be a 'hierarchy' of possible initial constituents: the subject pronoun, then *moult*, then *bien*. It is possible therefore that *moult* and *bien* are not fronted for pragmatic reasons, but simply to fulfill the syntactic requirement that something must precede the finite verb. [49]

49. *Moult* and *bien* fronting shows parallels to the phenomenon of stylistic fronting described in the generative literature (for MedFr see Roberts, 1993; Mathieu, 2006). Stylistic fronting (SF) is first described in Scandinavian by Maling (1980). In SF, the fronted constituent has no pragmatic interpretation, requires a null subject, and follows a strict Accessibility Hierarchy in terms of the type of constituent targeted (adverbs > adjectives > non-finite verbs) (Holmberg, 2000: 448–50). In this respect, such an analysis is a good fit to

3.4.2 Pragmatic properties

To demonstrate that a purely syntactic analysis of *moult* and *bien* fronting is correct, I will consider whether or not it has a pragmatic role (as argued, for example, by Buridant, 1992; Marchello-Nizia, 2006), focusing on *moult* fronting in *LaisMarie*. Marchello-Nizia's (2006: 161) data shows that pre-verbal *moult* is three times more common than post-verbal *moult* in Marie de France's *Lais*. If *moult* fronting is pragmatic in origin, a marked discourse reading is expected. For some tokens, this seems plausible:

(93) ore est Lanval mut entrepris,
mut est dolent e mut pensis. (*LaisMarie*, ll. 33–34)

For this example, a topic reading of the second *mut* seems possible: it reprises the pre-adjectival intensifier from the previous clause. Equally, the triple repetition of *moult* could be taken as emphatic, highlighting the misery of the impoverished Lanval. However, in other contexts, it is difficult to ascertain whether *moult* fronting has any particular semantic or pragmatic role.

(94) "Lanval," fet ele, "bien le quit,
vus n'amez gueres cel delit;
asez le m'ad hum dit sovent
que des femmez n'avez talent.
Vallez avez bien afeitiez,
ensemble od eus vus deduiez.
Vileins cuarz, mauveis failliz,
mut est mi sires maubailliz
que pres de lui vus ad suffert;
mun escient que deus en pert!" (*LaisMarie*, ll. 277–86)

One could argue that *mut* is fronted here, rather than the subject *mi sires*, for emphatic effect: it follows on from a context in which the queen is insulting Lanval's honour, and the fronting of an intensifier rather than the subject could be to emphasize the severity of the accusation. Yet there is little in the context to indicate such a reading unambiguously. However, there is a clear pragmatic reading in a similar example where the nominal subject is fronted:

---

the MedFr data: *moult* must be fronted in preference to *bien* or the predicative adjective, for example, and fronting cannot take place with an overt subject pronoun. *Moult* and *bien* fronting is possible with NP subjects, but SF only requires that Spec,TP (here equivalent to Spec,IP) be empty (Holmberg, 2000; Mathieu, 2006), and Vance (1997: 70–74) argues that NP subjects may remain in the VP. Thus, in MedFr, NP subjects are not predicted to block SF.

(95) Lanval defent la deshonur
e la hunte de sun seignur
de mot en mot, si cum il dist,
que la reine ne requist;
mes de ceo dunt il ot parlé
reconut il la verité,
de l'amur dunt il se vanta;
dolent en est, perdue l'a.
De ceo lur dit qu'il en ferat
quanque la curt esgarderat.
Li reis fu mut vers lui irez
tuz ses hummes ad enveiez
pur dire dreit qu'il en deit faire,
que hum ne li puisse a mal retraire.

(*LaisMarie*, ll. 371–84)

In (95), the initial subject has a marked topic reading, as the focus of the narrative switches from Lanval to the king. In her analysis of the *Chanson de Roland*, Marchello-Nizia makes an identical observation:

> Le sujet antéposé est certes systématiquement thématisé, mais en même temps emphatisé : il marque l'entrée en scène de l'un ou l'autre des protagonistes essentiels.
>
> (Marchello-Nizia, 1995: 96)

The same is not true of the post-verbal subject in (94), where the king is mentioned in passing during the queen's diatribe against Lanval. Constructions of this type provide good evidence that SVO word order is still realized in the CP (where the initial element is pragmatically marked) rather than the IP in the 12th century.

It could still be the case that initial *moult* is emphatically marked in (94). Yet other tokens in the text suggest that this is unlikely. *Moult* is commonly used with the attributive predicates *estre dolent* and *estre liés*. Of the six tokens of *estre dolent* in this text, all contain null subjects, and five show *moult* fronting, e.g.:

(96) hume estrange descunseillez
mut est **dolent** en autre tere,
quant il ne seit u sucurs quere. (*LaisMarie*, ll. 36–38)

The remaining token does not contain *moult*, and in this case the attributive adjective is fronted:

(97) mes de ceo dunt il ot parlé
reconut il la verité,
de l'amur dunt il se vanta;
dolent en est, perdue l'a. (*LaisMarie*, ll. 375–78)

My reading of this example does not suggest that *dolent* is pragmatically marked here. However, in *GormIsem*, a clearer pattern emerges. There are four tokens of *estre dolent*, all with null subjects, two with fronted *moult* and two almost identical tokens with a fronted adjective:

(98) lié serrunt cil k'aweron
dolent serrunt paien felun (*GormIsem*, ll. 277–78)

The fronting of the adjective is clearly for contrastive focus with *lié*.

In short, there is often a clear pragmatic reading of tokens of *estre dolent* from these two texts where an element other than *moult* is fronted. This is not the case when *moult* is initial. Moreover, *moult* is by far the most common initial constituent in tokens of *estre dolent*. From this evidence, I suggest that *moult* is used in pre-verbal position primarily as a pragmatically neutral way to fulfill the V2 requirement. It is possible that it is associated particularly with adjectival predicates for semantic reasons.

3.4.3 The decline of *moult* and *bien* fronting

Moult and *bien* fronting is significantly less common in texts dating from after 1250 (cf. table 6). Marchello-Nizia's (2006: 161–62) data show an increase in post-verbal *moult* at this time: in fact, post-verbal position is generally more common than pre-verbal position in most texts from the 13th century.

As *moult* and *bien* fronting cannot take place with an expressed subject pronoun, the most obvious explanation for this change is the increase in the use of the subject pronoun as shown in table 5. In turn, a widely accepted explanation for the increase in the use of the subject pronoun is that it is the 'most common' topic in a V2 system (e.g. Vennemann, 1974; Harris, 1978). Yet in this case the explanation is problematic. Firstly, since *moult* and *bien* are not topics, it is not clear that the V2 system of the 12th century is intrinsically topic-initial. Secondly, *moult* and particularly *bien* fronting may be replaced by the use of an expletive pronoun:

(99) del duel ne convient mie a parler qui illuec fu faiz que onques plus grant ne fu faiz por home et **il le dut bien estre** car onques hom de son aage ne fu plus amés de ses homes et de l'autre gent (*ConqVilleh*, ¶37)

Compare a similar construction from *Charrete*, where *bien* is fronted:

(100) **bien** puet estre (*Charrete*, l. 790)

Expletive pronouns cannot be descended from topic-initial structures, since they have no referent. Pragmatically, one unmarked word order is simply replaced by another.

However, the development is predicted using the prosodic Stress Constraint. Following the emergence of group stress, unstressed and unemphatic fronted *moult* and *bien* are incompatible with the Stress Constraint. Unlike the subject pronoun, it is difficult to see from the syntactic data how these adverbs could be reanalysed in a position other than Spec,CP. Instead, the construction was increasingly avoided in favour of SpVAdv order, which does not contravene the Stress Constraint.

Tokens of *moult* and *bien* fronting do not disappear, and are still attested in 14$^{\text{th}}$-century texts:

(101) li quins sages ot nom Ruthonius et en seurnom Philopater qui vault en rommans comme 'pere d'amour', et **bien** fu drois qu'il eüst tel seurnom (*Berinus*, §6)

(102) **molt est** malicïeus et sage
l'anemi de l'umain lignage (*ComteAnjou*, ll. 243–44)

There is some evidence from the texts written by Christine de Pisan from the early 15$^{\text{th}}$ century that these structures have been reanalysed. Firstly, *moult* and *bien* occur as the first element in Type II V3 word orders. The syntax of this construction is novel in two ways: the adverb is not directly pre-verbal, and it is fronted despite the presence of the subject pronoun:

(103) Se m'amour voulsisse ottroyer
ja pieça m'a esté requise,
mais j'ay ailleurs m'entente mise.
On vendroit trop tart au proyer,
et pour tant bien je vous avise
se m'amour voulsisse ottroyer. (*Christine*, rondeau LVI, ll. 1–6)

(104) Et ainsi dura ceste pestillence tant que le ·Ve· filz vint, lequel par vertu de saige conseil attray lumiere tellement que les tenebres de cele mortel secheresse furent chaciez; par laquelle voie ot congnoissance de dois, et sourse d'eaue vive a grant habondance si que lui meismes ja tout perdus et sechiez en fu arrousez et vivifiez. Et tant pourchaça par saiges maistres que l'abondance du fleuve vivant rendi ruissiaux,

sourses et fontaines en si grant quantité que toutes mes plantes en furent arrousees et vivifiees. O! noble cultiveur fu ycellui, **car moult il augmenta** et percrut la perpetuité de mon heritaige en tel maniere qu'il fut le premier registre de ma salvable gloire.

(*AdvisionChr*, I, §8, ll. 13–22)

I reproduce these examples in context to illustrate a second important change: the fronted adverb is now clearly emphatic. Example (103) is a *rondeau* and the lady's reply to a would-be suitor: the fronting of *bien* is contrastive, emphasizing the lady's advice for her suitor to overcome the normal reticence to come to prayer. In (104), the narrator is a princess and an allegorical representation of the French nation. Having enumerated the greatness of the king's achievements in the previous sentence, the use of *moult* here places clear emphasis on the magnitude of their effect on his country.

An emphatic reading of the fronted adverb is not restricted to the verb-third structure. In the following example from *MutFortune*, *moult* may be read with emphatic focus. The author is discussing precious stones, drawn from Fortune's fountain:

(105) il n'est escharboucle, sanz faille,
ne riche rubis qui les vaille;
en l'une avoit tant de vertus
que, puis le temps le roy Artus,
n'ot plus riche roy, n'emperiere;
resplendissant de grant maniere
fu et plus luisant que chandoile,
et de couleur sembla estoile;
belle fu et tres gracieuse
et ancore plus precieuse,
mais **moult fu penible a trouver.** (*MutFortune*, ll. 231–41)

Following on from a long description of the virtues and beauty of the stone, *moult* fronting here emphasizes the single drawback: the difficulty of finding it.

I assume that the pragmatic reading here is associated with an emphatic focal stress. Thus, the 15$^{\text{th}}$-century tokens of *moult* and *bien* fronting once again conform to the Stress Constraint, but in order to do so acquire a pragmatically marked reading not present at an earlier stage of the language.

The earliest attestation of Type II V3 with *moult* or *bien* is found in *PassPalat*, from the early 14$^{\text{th}}$ century:

(106) a Pilate nous le merrons
et sus lui metrons tel tort,
bien il deservira la mort. (*PassPalat*, ll. 312–14)

The manuscript dates from the first half of the 14th century and the editor has made no emendations to this line, so there is no reason to suppose that the token is not genuine. CSV word order is not atypical here, occurring also in line 312 and, in the manuscript, in line 313 (*nous metrons*). If the reanalysis of *moult* and *bien* fronting is triggered by the loss of word stress in the mid-13th century, tokens of the new word order from the early 14th century are predicted. Further research on a wider corpus would be needed to establish how common such tokens are in the 14th century.

To summarize, I propose that the development of *moult* and *bien* fronting may be explained by the Stress Constraint proposed in section 3.2.2. In the 12th century, the fronting of these short adverbs was an unemphatic strategy serving simply to satisfy the V2 constraint in certain null subject clauses. The word stress on these adverbs was sufficient to conform to the Stress Constraint. Following the loss of word stress, this unemphatic fronting of light adverbs become disfavoured. Where no emphasis was required, SpVC word order became more common in the 13th century. In accordance with the Stress Constraint, initial *moult* and *bien* were reanalysed as emphatic, acquiring both an emphatic stress and a marked pragmatic reading.

3.5 *Or*

Unlike *moult* and *bien*, the data presented in table 6 show that the adverb *or* is frequent in initial position throughout the MedFr period. Given that I have argued that the Stress Constraint causes a decline in the frequency of pre-verbal *moult* and *bien*, it must be shown that *or* is not predicted to be affected in the same way.

I suggest that *or* is not affected by the Stress Constraint because it is not a fronted constituent, but instead is directly merged in the CP. Similar analyses have been proposed for *si* (Ferraresi and Goldbach, 2002; Poletto, 2005; Ledgeway, 2008) to which I return in section 3.6. If this is the case, the emergence of group stress would cause stress on initial *or* to be lost, but without any violation of the Stress Constraint.

The first argument that *or* is not fronted from within the clause is a quantitative one. In the early-8, base and prose subcorpora combined, only five tokens of post-verbal *or* are attested, as against 77 in initial position in verb-second

clauses.

(107) judas ne dort **ore** noient (*PassJongl*, l. 535)

(108) sire, fait li cevaliers du pont, il ne me plaist que il i passe ne il n'i passera **ore** mie

(*TristanPr*, §9, l. 12)

Moreover, unlike fronted *moult* and *bien*, there is a clear pragmatic difference between tokens of pre-verbal *or* and those of post-verbal *or*. For van Reenan and Schøsler (1995, 2000), *or* is a marker of 'topic switch', which they define as 'a change of subject [...] from one main clause to the other.' (2000: 63). Buridant makes a similar claim:

> Fondamentalement, il [*or*] fait référence au temps T° de son énonciation en signalant, par cette référence, l'intention du sujet d'effectuer la rupture avec ce qui précède pour développer une nouvelle assertion, et marque la confrontation de deux actes d'énonciation à des fins conclusives.
>
> (Buridant, 2000: §417)

This is not true of the tokens of post-verbal *or* above: for example, in (108) the subject of the clause containing *ore* is identical to that of the previous clause. Post-verbal *or* instead retains the core semantic value of 'at this time', with no additional pragmatic features.[50] A third property which demonstrates that pre-verbal *or* has a different status from other fronted constituents is that it occurs before imperative verbs:

(109) **or** asëez! (*Brendan*, l. 452)

(110) **or** regardez donques, madame, ce povre qui oncques ne le vist, ne le cognoust, ne parla a luy (*JehSaintre*, I, p. 38)

The standard analysis of V1 constructions such as imperatives in V2 languages is that the specifier position is filled by an 'empty' discourse operator (cf. Vance, 1997: 18–20), leaving the verb in initial position. If Spec,CP is filled by an empty operator, there is no cause for a constituent to be fronted to fulfill the V2 requirement.

Since pre-verbal *or* is not affected by the Stress Constraint, there is no motivation for a reanalysis following the emergence of group stress. Tokens of *or*–SV word order are not attested in the full corpus, while V2 tokens of CVSp are still attested in the 14$^{\text{th}}$ and 15$^{\text{th}}$ centuries:

50. These pragmatic, semantic and positional differences may also be connected with a morphological distinction, since all the post-verbal tokens examined are disyllabic. The regularity of this distinction would need to be verified over a wider corpus. In the present section I follow scholars such as Buridant (2000: §§417–21) in treating *or* and *ore(s)* as equivalent.

(111) **or** vous eusse-je bien mors se je vousisse (*VieLouis*, ¶40)

(112) **or** fus je, la Dieu mercis, si prochaine privee amie de icelle que, de sa grace, faveur ot a me descouvrir les secrez de son cuer
(*AdvisionChr*, I, §5, ll. 9–11)

(113) **or**, sire, vous ay je cy (*JehSaintre*, I, p. 35)

However, example (113) clearly demonstrates that by the mid-15th century, a dislocated reading of initial *or* is possible. Here, it precedes a vocative, and is clearly not within the PhPh of the finite verb. Considering data from the full corpus, tokens of dislocated *or* are found in 15th-century texts, particularly in direct speech or in plays. One such token is attested in *Testament* and *PassGreban*, and three in *Pathelin*, e.g.:

(114) **or**, puisque chascun enlumine,
a tant nostre quart jour termine. (*PassGreban*, ll. 296–97)

(115) **or** vrayement j'en suis attrappé
car je n'avoye intencion
d'avoir drap, par la passion
de nostre seigneur, quant je vins. (*Pathelin*, ll. 194–97)

These tokens would seem to indicate a change in the syntax of *or* in the latter part of the 15th century. It becomes a clause-initial adverb, able to occupy a higher position in the C-domain than Spec,CP. In *Jehan de Saintre*, Vance (1997) notes a token of CSV with *or*, again with a vocative between *or* and the main clause:

(116) **Ores**, roy d'armes, je vous prie
(*Jehan de Saintre*, p. 102, l. 13; Vance, 1997: 265)

Vance (1995b: 186) notes that this text 'appears to be slightly anomalous in preferring non-inversion [of the subject] to a greater extent than the other Middle French texts', and estimates that only 40% of total subjects, and only 24% of subject pronouns, are post-verbal following an initial non-subject constituent. Such a low rate of subject pronoun inversion suggests that a general syntactic reanalysis has taken place, at least in the grammar of the author: the verb can no longer productively move to the C-domain, and inverted structures are only retained as relics. Moreover, it is perhaps significant that all the tokens we have seen of *or* in this new configuration are drawn from direct discourse. In part, this may be because *or* is anyway more frequent in direct discourse (van Reenan and Schøsler, 2000: 87); alternatively, it may indicate that the V2 constraint has all but disappeared from the spoken language of the late 15th

century while still being retained in written narrative. In any case, the much later date of the change in the syntax of *or* leads me to suggest that in this case reanalysis is triggered by a general syntactic change: the loss of V-to-C movement (Vance, 1995b).

To summarize, the Stress Constraint has no effect on the syntax of *or* as it is base-generated in the CP rather than moved from within the clause. Changes in its syntax in the late 15th century are not due to prosodic change but to more general syntactic changes in the language.

3.6 A brief note on *si*

The analysis we have proposed for *or* can easily be extended to the most common pre-verbal adverb, *si*, which shows many similar syntactic and pragmatic properties. *Si* is analysed in a number of studies, both specific to MedFr (Marchello-Nizia, 1985; Fleischman, 1991, 1992; van Reenan and Schøsler, 1992, 1993, 1995, 2000; Ferraresi and Goldbach, 2002) and in a wider Romance perspective (Poletto, 2005; Ledgeway, 2008), and thus I propose to treat it only very briefly to show that it is compatible with the present analysis.

There are strong similarities between *si* and *or*. Firstly, pre-verbal *si* cannot be separated from the finite verb (except of course by negation and object clitics) (Skårup, 1975: 290; Marchello-Nizia, 1985: 163; Fleischman, 1991: 268; Ferraresi and Goldbach, 2002: 11). Secondly, *si* has a clear discursive function. For Marchello-Nizia (1985: 2–3), *si* has a fundamentally assertive value, confirming the truth of the proposition. Fleischman (1991: 256–61), on the other hand, argues that *si* functions primarily as a topic continuity marker, in opposition to the subject pronouns which introduce a new topic, an analysis also adopted by van Reenan and Schøsler (1992, 1993, 1995, 2000). Particularly relevant to this analysis is the fact that *si* is never attested at the beginning of 'the opening sentence of a thematic paragraph' (Fleischman, 1991: 448; cf. also Marchello-Nizia, 1985: 30; Lemieux and Dupuis, 1995: 96). Thirdly, post-verbal *si* is rare, and lacks a pragmatic or discursive function aside from its core semantic meaning 'thus, in this way' (Marchello-Nizia, 1985: 14). Consequently, as with pre-verbal *or*, we may suppose that pre-verbal *si* is merged in the lowest specifier of the CP rather than fronted to this position, and is therefore exempt from the Stress Constraint. Precisely this analysis is proposed by Poletto (2005) for Old Italian, a similar proposal is made for MedFr by Ferraresi and Goldbach (2002) and for Old Neapolitan by Ledgeway (2008).[51]

51. The proposals of Ferraresi and Goldbach (2002) and Ledgeway (2008) argue that *si* is

3.7 Summary

In this section, the prosody of the initial constituent in main clauses was examined, with particular attention paid to the subject pronouns and the four most common adverbs in the 12th and 13th centuries. These were argued to be 'light' initial elements, which do not form their own PhPh, and thus are predicted to become unstressed following the emergence of group stress. I have argued that this loss of stress in the 13th century caused structures with fronted light elements to violate a prosodic constraint, requiring constituents fronted to the CP to be stressed. In the case of the subject pronouns, this favoured the reanalysis of CP-SVO word order, still attested in the 12th century, as IP-SVO order. This reanalysis in turn triggered the introduction of Type II V3 word orders in the 14th century. In the case of *moult* and *bien*, no such reanalysis was available, and adverb fronting became less common. Moreover, by the 15th century, there is evidence that whether in CV or CSpV word order, the remaining pre-verbal tokens of *moult* and *bien* had been reanalysed as stressed and emphatic. Thus fronting to CP was once again compatible with the Stress Constraint. *Or* and *si*, on the other hand, were not fronted to Spec,CP but instead base-generated there, and so were unaffected by the Stress Constraint. V2 word orders with *or* and *si* thus remained stable until the V2 constraint disappeared in the late 15th century.

If this analysis is correct, it demonstrates that the emergence of group stress had important consequences for the verb-second system of MedFr. Firstly, as Vance (1997) hypothesizes, it led to the introduction of CSpV word order through favouring the reanalysis of SVO word order as IP-internal. Secondly, it led to a decline in the number of light pre-verbal elements being fronted to Spec,CP from within the clause (e.g. *moult* and *bien*) as these could no longer be realized as stressed in this position. There was thus less evidence for one of the key features of a V2 system: the fronting of any clause-internal element to the initial position. The adverbs *or* and *si* were unaffected by the emergence of group stress, remaining as unstressed initial elements in CVSp word order until the V2 constraint was lost in the late 15th century.

Prosodic change does not eliminate V2 — V2 structures are not infrequent

a functional head adjoined to the finite verb in C^0. Crucial in both analyses is that *si* may regularly occur in verb-third orders with constituents usually assumed to occur in the lowest specifier of the CP, such as pre-verbal objects (Ferraresi and Goldbach, 2002: 16; Ledgeway, 2008: 449). This property is unique to *si* of all pre-verbal elements. Following Roberts (2004), constituents merged in a C^0 may satisfy the V2 constraint. However, whether *si* is merged as a head or as an adverbial specifier, we predict that it will become unstressed following the emergence of group stress without infringing the Stress Constraint.

in the 14th century and early 15th century — but it causes some constructions with light initial constituents to be susceptible to reanalysis as non-V2 structures.

Chapter summary

This chapter set out to investigate the mechanism and consequences of prosodic change. In section 1, it was shown that the stress patterns generated by group-stress and word-stress grammars for the PhPh containing the head noun of the noun phrase were very similar in MedFr. Such an overlap provided the conditions for a reanalysis of the stress rules. Moreover, conditions were more favourable to reanalysis in the pre-1250 period, a finding which corroborates the proposed chronology of prosodic change suggested in chapter three. In section 2, I showed that post-verbal pronouns in PhPh-final position were stressed from the early 12th century. I argued that these stressed forms generalized as the result of a rule requiring the PhPh to end in a PWd. This rule was shown to be compatible with a word-stress grammar, but its effect was to produce regular PhPh-final stress, a further condition necessary for the emergence of group stress. From these findings, it is clear that an appropriate context for the reanalysis of the stress rules existed in the 12th century.

In section 3, I argued that the emergence of group stress had important consequences for the V2 system of MedFr. Short clause-initial constituents do not form their own PhPh, and thus became unstressed following the loss of word stress. However, I suggested that stress is a fundamental property of constituents fronted to clause-initial position, and thus the loss of stress triggered a syntactic reanalysis of these structures, notably subject-initial clauses, which was not compatible with the V2 constraint. The existence of non-V2 structures, particularly the CSpV construction, contributed to the eventual loss of V2 in the late 15th century.

Summary and Conclusions

The present thesis has investigated the development of group stress in French over the medieval period. In particular, I have sought to clarify four key aspects of this prosodic change: its character, chronology, mechanism and consequences.

1 Corpus and inter-textual differences

Central to tackling the studies in this thesis was an electronic corpus of 87 extracts from medieval texts with extensive machine-processable annotation. The metrical annotation of the corpus, not found in any other corpus of French, was used in the wide-ranging quantitative studies of versification in chapter three. Moreover, since the corpus combines both syntactic and prosodic annotation, quantitative studies of phonological phrasing could be carried out. In any language, this would enable a more sophisticated account of stress-related phenomena; in French, where the phonological phrase becomes the primary stress-bearing unit, it is fundamental. In the present thesis, this unique combination of metrical, syllabic and syntactic annotation is used in two studies in particular: the study of the position of prosodic constituent edges within lines of verse in chapter three (study 3), and the study of patterns of prosodic words within phonological phrases headed by a noun in chapter four (§1). Studies of this kind using corpus-derived quantitative data have never been previously attempted in French.

Dividing the corpus into subcorpora of comparable texts from different time periods enabled change over time to be studied in different verse forms, dialects and text types. In order to carry out accurate diachronic and diatopic studies of rhythmic change, particular attention was paid to the reliability of the dating and localization of the texts used in the corpus.

In chapter three (study 2), the subcorpora were used to show that narrative, lyric and theatrical texts show different rhythmic properties. Strong

rhythmic organization, in particular a tendency to stress the fourth syllable in the octosyllabic line, was shown to be a feature of early narrative texts. Lyric and theatrical texts (apart from the *Jeu d'Adam*) do not show similar rhythmic organization at any point. Data from the wide range of octosyllabic narrative texts from the 12th and 13th centuries included in the corpus suggest that it is those narratives with the strongest links to spoken oral performance that show the strongest rhythmic organization. This prosodic finding complements existing discourse-based analyses (e.g. Zumthor, 1972; Marnette, 1998) in highlighting the importance of the performance context of medieval texts. Dialectal factors were found to have no effect on the rhythm of octosyllabic verse.

2 Group stress and linguistic change

The aspect of medieval prosodic change on which this thesis has focused is the emergence of group stress. The character of the change was investigated through a study of the properties of the ModFr group-stress system. An inherent property of this system was argued to be 'stress deafness' (cf. chapter one, §2). Since the opposition between stressed and unstressed syllables cannot be phonologically distinctive, experimental studies show that speakers experience difficulty in processing minimal pairs of stimuli distinguished only by the position of stress. In chapter one (§1), we saw that the notion of group stress can be formalized by assuming that stress is assigned to the phonological phrase, a prosodic unit defined in part by syntactic and rhythmic factors but subject to some variation in its realization (cf. chapter one, §1). Primary stress is consistently realized on the final syllable of the phonological phrase. No stress is consistently realized on prosodic words within the phonological phrase. Indeed, so variable is phonological phrase-internal stress that it was argued to be primarily rhythmic, and not associated with any particular prosodic word. The lack of stress on prosodic words within the phonological phrase differentiates a group-stress grammar from a word-stress grammar in which each prosodic word has an underlying primary (and possibly secondary) stressed syllable. In a word-stress system, phenomena dependent on stress position, such as vowel lengthening and vowel reduction or deletion, apply consistently to all prosodic words. Since such processes are well-attested in the pre-textual history of French, we may conclude that the group stress of modern French developed from a word-stress system.

In order to date the emergence of group stress, we have focused only on those linguistic developments observable in the textual record which are clearly stress-conditioned. The *terminus ad quem* for the change was provided by Palsgrave's description of French accentuation from 1530, which was taken as evidence for the existence of a group-stress system at this time. A *terminus a quo* was more difficult to establish. A regular word-stress rule, which developed as a consequence of the deletion or reduction to schwa of unstressed vowels, was a necessary condition for the emergence of group stress. While it is clear that the processes giving rise to a regular stress rule began pre-textually, it is less certain at what point they had run to completion. However, evidence for the oxytonic pronunciation of Latin in the *Passion de Clermont* from the early 11$^{\text{th}}$ century strongly suggests that a regular word-stress rule was active in French at this time. Phonological changes deriving from stress-conditioned allophony, such as the diphthongization of primary stressed vowels, are in progress by the time of the *Sequence of Saint Eulalia* in the late 9$^{\text{th}}$ century, and are no longer allophonic variants by the 12$^{\text{th}}$ century. On these pieces of evidence, the *terminus a quo* for the emergence of group stress was set before the 11$^{\text{th}}$ century.

Fixing a date more precisely within the five centuries between 1000 and 1500 marked one of the significant contributions of the thesis. The major piece of evidence used to date the change was derived from patterns of versification. In chapter one (§4), I suggested that syllabo-tonic versification, in which the position of word stress within a line of verse is constrained, cannot occur in a language whose speakers manifest stress deafness effects, since the difference between stressed and unstressed syllables is not salient. Previous studies show that the earliest octosyllabic verse texts from the 11$^{\text{th}}$ century and early 12$^{\text{th}}$ century tended to alternate stressed and unstressed syllables, a tendency much less clearly observed in verse from the late 12$^{\text{th}}$ century. In chapter three, an extensive study of the rhythm of verse was carried out, with important findings from the point of view of the chronology of prosodic change:

(I) All narrative octosyllabic verse texts studied from before the mid-12$^{\text{th}}$ century show a significant tendency for the fourth position in the line (and to a lesser extent the second and the sixth) to contain a 'prominent' syllable. Similar texts from after the mid-13$^{\text{th}}$ century show no such tendency (chapter three, study 1).

As I have argued that stress deafness and the rhythmic organization of word stress are incompatible, finding (I) suggests that stress deafness cannot have

emerged before the mid-12th century. This represents a much later *terminus a quo* than was provided by segmental phonological evidence. Moreover, the disappearance of rhythmic organization from texts by the mid-13th century could be taken to indicate that stress deafness, and therefore group stress, has emerged by this point. Two further findings, however, suggest that factors other than the organization of word stress may have contributed to this change:

(II) No evidence is found for strong rhythmic organization in lyric and theatrical texts. Moreover, even within narrative, the context of performance of the text plays an important role. Texts most strongly linked to spoken oral performance are most likely to be rhythmically organized.

(III) The fourth syllable in octosyllabic narrative texts from before the mid-12th century is usually associated with the edge of a phonological phrase, similar to a cæsura in ten- and twelve-syllable verse. Similar texts from after the mid-13th century do not favour phonological phrase edges after any particular syllable in the line (chapter three, study 3).

From finding (II), it could be argued that the decline in rhythmic organization in octosyllabic verse does not indicate linguistic change at all, but instead reflects the increased proportion of clerical texts intended for private reading. From finding (III), it could be argued that French versification never favours a syllabo-tonic style organization of word stress, but instead that early octosyllables show a constraint on the placement of a phonological phrase boundary, a restriction which is compatible with a group stress system. Yet evidence for organization of word stress is strong in a number of texts from before 1250, in particular the *Passion de Clermont*, *Gormont et Isembart* and Gautier de Coinci's *Miracles de Nostre Dame*. In these texts, there was a tendency to stress the fourth syllable of the line even if this involved shifting the phonological phrase boundary to the fifth syllable to accommodate a post-tonic schwa. The mid-line phonological phrase edge was therefore shifted to the fifth syllable, resulting in a regular stress pattern (4+4) but an unequal syllabic division of the line (5+3) (chapter three, study 3). Moreover, while the disappearance of strong rhythmic organization of this type may have been due to the changing context of performance of octosyllabic texts, the existence of rhythmically organized texts from the 12th century and the early 13th century allowed us to conclude that group stress cannot have fully developed in French before the mid-13th century. The emergence of group stress was thus dated between the mid-12th century and the mid-13th century.

The mechanism by which group stress emerged was examined in chapter

four (§1). From the point at which a fixed word-stress rule developed, every prosodic word ended in a stressed syllable, or a stressed syllable followed by a post-tonic schwa. Moreover, the development of stressed pronouns in post-verbal position demonstrates that every phonological phrase had to end in a prosodic word, a development first attested in the 11th-century *Vie de saint Alexis* (cf. chapter four, §2). Thus, every phonological phrase ended with a primary stress. Based on an examination of phonological phrases containing the head noun of the noun phrase and one other prosodic word, I investigated the position of the word stress on the non-final prosodic word within the phonological phrase. The following key findings were obtained:

(I) In 87.5% of tokens, the non-final word stress either fell on the first syllable of the first content word in the phonological phrase or clashed with a following phrase-final stress (or, in some cases, both).

(II) In 22.4% of tokens, the non-final word stress clashed with the phrase-final stress.

Finding (I) shows that nearly 90% of phonological phrases containing the head noun of the noun phrase and one other prosodic word could be analysed in two ways. Firstly, they could be generated by a word-stress grammar with a clash resolution rule that deletes the first of two clashing stresses. Secondly, they could be generated by a group-stress grammar that places a secondary stress on the first syllable of the first content word in the phonological phrase. There was thus sufficient ambiguity in the output for a reanalysis of the stress rules to take place. Moreover, finding (II) shows that a significant proportion of prosodic words would have lost their primary stress due to a clash resolution rule, a fact which may have favoured the spread of a group-stress analysis. From a chronological point of view, the evidence from versification is strengthened by the fact that conditions were more favourable for reanalysis in the pre-1250 period to a statistically significant degree. In this period, 91.2% of tokens were compatible with either analysis (cf. finding I), while 25.1% of tokens show clash resolution (cf. finding II).

Finally, I have argued that the emergence of group stress contributed to the decline of the verb-second grammar of medieval French. While rejecting the theory that verb-second is caused by a prosodically strong clause-initial position (chapter one, §3.2), I claimed that the loss of primary stress in all but phonological phrase-final position gave rise to a change in the prosody and subsequently the syntax of short clause-initial constituents (chapter four, §3). Using evidence from versification, I demonstrated that these monosyllabic

constituents did not form their own phonological phrase, but instead constituted the first element in a longer phonological phrase which also contained the finite verb. From the 13th century, following the development of a group-stress grammar, these phrase-initial constituents became unstressed. Working within a broadly generative model of clause structure, I suggested that there is an inherent tension between a lack of stress and the movement of a constituent to clause-initial position from within the clause. Over the course of the 13th and 14th centuries, two developments in word order caused these unstressed monosyllabic constituents moved to the initial position of a V2 main clause (i.e. Spec,CP) to be eliminated:

(a) Reanalysis of subject pronoun–verb word order as IP-SVO (non-V2) rather than CP-SVO (V2). Emergence of CSV word order from the early 14th century.

(b) Decline in the frequency of pragmatically unmarked *moult* and *bien* fronting. Reanalysis of fronted *moult* and *bien* as pragmatically marked and stressed initial constituents.

The theory also predicted the lack of effect that prosodic change had on the syntax of *or* and *si*, unemphatic pragmatically-marked adverbs which were base-generated in the Spec,CP position rather than fronted from within the clause. Crucial to this analysis was the assumption that while prosodic requirements do not trigger syntactic movement, certain kinds of syntactic movement, in this case the obligatory movement of clause-internal constituents to initial position to fulfill the V2 requirement, may be associated with prosodic marking. The loss of this prosodic marking due to wider linguistic change therefore caused a reanalysis of the underlying syntactic structure. If this analysis is correct, it suggests that prosodic change may contribute to syntactic change.

3 Further questions

One of the most interesting observations from the study of versification is the apparent association of rhythmic organization with orally performed narrative. The idea that verse form is associated with the performance context of the text is not new: the emergence of continuous verse in the mid-12th century, for example, is linked a change in the 'mode d'audition' (Zumthor, 1972: 340). However, the idea that rhythmic organization *within* the line of verse is closely associated with performance context has not been systematically studied. As differences between narrative texts were not modelled using different subcor-

pora, the conclusions reached in this study are only provisional, especially in view of the debate as to the extent to which romances and *fabliaux* were orally performed. It would be interesting to examine a corpus of *fabliaux* and romances from the 13th century to examine whether the rhythmic difference can be demonstrated more rigorously, and whether on a text-by-text basis it can be correlated with more discursive marks of orality.

In chapter four, I presented short analyses of two phenomena which have long been considered by historical linguists to be affected by prosodic change: the development of pronominal forms and the loss of V2 effects. In both cases, referring to the prosodic hierarchy provided a better understanding of the role of stress in these developments. With reference to post-verbal pronouns, I argued that the requirement for a phonological phrase to end with a prosodic word caused a particular pronominal allomorph to be selected. This approach is crucially different to previous analyses of the relationship between stress and morphological form discussed in chapter one (§3.2.2), which assume that stress affected morphological form through productive phonological rules. The view adopted in the present thesis may therefore shed new light on other aspects of the development of function words in medieval French. In the case of verb-second, I argued that the syntactic movement associated with V2 is usually associated with a stress in the surface form. Clearly the predictions of the Stress Constraint must be verified both through a study of other initial constituents and data from subordinate clauses. Moreover, it would be interesting to consider French in a wider Romance perspective. For example, all medieval Romance varieties show V2 to some extent (cf. Benincà, 2006), and the system is lost in all cases. Since only French shows prosodic change, we would therefore expect the disappearance of V2 to follow a different pattern in other Romance dialects. Italo-Romance varieties, for example, show a much higher proportion of Type I V3 (Benincà, 2006) throughout the time period. It is also unclear that the Type II V3 order, which I have suggested emerges as a prosodically-induced reanalysis of the subject pronoun, could have developed in southern Italo-Romance or Ibero-Romance, where subject pronouns remain emphatic to this day.

Lastly, a crucial question which has not been examined is *why* group stress emerges in French. Certainly, we have seen that a number of necessary conditions are satisfied in the pre-1250 period: a fixed word-stress rule, a requirement that every phonological phrase end in a prosodic word, and considerable overlap between the stress patterns predicted by word- and group-stress grammars

(chapter four, §1). However, all that this demonstrates is that either analysis of the data is possible.

One possible explanation is that a group-stress grammar is innately preferred to a word-stress grammar. Peperkamp (2004) argues that only utterance-level stress regularities are salient to pre-lexical infants, and if stress is regular at the utterance level, it will not be encoded in the lexicon. Thus, we may speculate that if infants' initial analysis of stress is based on utterance edges, it is perhaps the case that a grammar in which stress is used only to mark the edges of prosodic constituents will be preferred.

A second possibility is that given the degree of overlap between the two grammars in the pre-1250 period, word-stress and group-stress grammars simply co-existed. It is perhaps not until the post-1250 period, when the proportion of tokens distinguishing the two analyses increases (perhaps due to the influx of longer Latinate words), that the two grammars would have produced more noticeably variant stress patterns. Once variants produced by word- and group-stress grammars exist as a form of perceptible synchronic linguistic variation, the reasons for the final emergence of group stress are perhaps as likely to be found in the wider socio-cultural context as they are in the internal structure of the language itself.

Bibliography

Corpus

Full details of each of the texts included in the corpus are provided in the following format:

Abbreviation: (Genre, Form). *Full title* and author

Composed: Date of composition, place of composition (for texts before 1300) / place of composition in Dees (1987) ('Dees': region and correspondance coefficient from Dees, 1987, cf. chapter two, note 27)

Manuscript: Manuscript reference [date of manuscript, place of composition of manuscript].

Source: Edition used (DEAF: Abbreviation used in *Dictionnaire étymologique de l'ancien français*)

Extract: Extract used

Key to abbreviations:

Genre

`ChG`	*Chanson de geste*
`Chr`	Chronicle
`Lyr`	Short lyric pieces
`RelNarr`	Religious narrative
`RelPlay`	Religious play
`Rom`	Romance
`SecPlay`	Secular play
`VieS`	Saint's life

Form

`8s`	Octosyllabic verse
`10s`	Ten-syllable verse
`12s`	Twelve-syllable verse
`L`	Short lyric form

`P`	Prose

Region

`C`	Central
`N`	Northern
`S`	Southern
`W`	Western
`E`	Eastern
`Eng`	England

Dialect

`Ch.`	Champagne
`Fr.-Com.`	Franche-Comté
`Fr.-Pr.`	Franco-Provençal
`IdeF`	Île-de-France
`Lorr.`	Lorraine
`Pic.`	Picardy
`unk.`	unknown

3Jugemens: (Dit, 10s). *Le Livre des trois jugemens*, by Christine de Pisan

Composed: 1400

Manuscript: BL Harley 4431 (A2) [*c.*1413, C (IdeF)]

Edition: Altmann (1998)

Extract: Le Livre des trois jugemens (pp. 153–202): ll. 1–477

AbrCheval: (Treatise, 8s). *Li Abrejance de l'ordre de chevalerie*, by Jean Priorat de Besançon

Composed: *c.*1290, E (Fr.-Com. / Fr.-Pr.)

Manuscript: BN fr. 1604 [*c.*1300, E (Dees: Fr.-Com., 86)]

Edition: Robert (1897) (DEAF: JPrioratR)

Extract: ll. 1–500

AdamHale: (Lyr, L). *Chansons*, by Adam de la Halle

Composed: *c.*1280, N (Arras)

Manuscript: BN fr. 847 (P) except BN fr. 12615 (T) (no. XV) [MS P last quart. C13, MS T 2nd quart. C14; MS P loc. not given, MS T N (Arras)]

Edition: Marshall (1971) (DEAF: AdHaleChansM)

Extract: chansons II, IV, VI, XV, XXI, XXVI, XXXIV–XXXVI

AdvisionChr: (Treatise, P). *L'Avision Christine*, by Christine de Pisan

Composed: 1405

Manuscript: ex–Philipps 128 (C) [autograph]

Edition: Reno and Dulac (2001)

Extract: première partie, chs I–X

Alexandre: (Rom, 12s). *Le Roman d'Alexandre*, by Alexandre de Paris

Composed: *c.*1185, C (IdeF)

Manuscript: BN fr. 25517 (G) [2nd half C13, N (Pic.)]

Edition: Armstrong (1937) (DEAF: AlexParA)

Extract: branch II, laisses 130–149 (ll. 2762–3100)

Alexis: (VieS, 10s). *La Vie de saint Alexis*

Composed: end C11, W (Norman)

Manuscript: Hildesheim St. Godehardi (L) [*c.*1120, Eng]

Edition: Storey (1968) (DEAF: AlexisS2)

Extract: ll. 1–400

AlexisA: (VieS, 12s). *La Vie de saint Alexis*

Composed: mid C14

Manuscript: BN fr. 1555 (A) [beg. C15]

Edition: Paris and Pannier (1887) (DEAF: AlexisQP)

Extract: Rédaction en quatrains alexandrins monorimés du XIVe siècle (pp. 329–88), quatrains 1–82

AlexisO: (VieS, 12s). *La Vie de saint Alexis*

Composed: beg. C13, unk.

Manuscript: Oxford Bodl. Canonici Misc. 74 (O) [beg. C13, NE (Dees: Walloon, 81)]

Edition: Stebbins (1971) (DEAF: AlexisAlOS)

Extract: ll. 1–333

AmiAmile: (ChG, 10s). *Ami et Amile*

Composed: *c.*1200, unk.

Manuscript: BN fr. 860 [2nd half C13, E (Lorraine, DEAF) / C (Dees: Haute-Marne, 80)]

Edition: Dembowski (1969) (DEAF: AmAmD)

Extract: laisses 1–22 (ll. 1–399)

BaratHaim: (Fabliau, 8s). *Barat et Haimet*, by Jean Bodel

Composed: *c.*1195, N (Pic.)

Manuscript: BN fr. 19152 (D) [end C13, C (IdeFr)]

Edition: Noomen and van den Boogaard (1983–98) (DEAF: HaimBarNo)

Extract: Barat et Haimet (in vol. 2: 27–76), full text (508 lines)

Behaingne: (Dit, 10s). *Le Jugement dou roy de Behaingne*, by Guillaume de Machaut

Composed: before 1342

Manuscript: BN fr. 1584 (A) [before 1377]

Edition: Palmer (1984) (DEAF: GuillMachBehP)

Extract: ll. 1–508 (incl. some 4-syll lines)

BelleDame: (LyrNarr, 8s). *La Belle dame sans mercy*, by Alain Chartier

Composed: 1424

Manuscript: Toulouse Bibl. mun. 826 (Qd) [early C15]

Edition: Laidlaw (1974)

Extract: La Belle Dame sans mercy (pp. 328–70), stz. I–LXV (ll. 1–504)

Berinus: (Rom, P). *Bérinus*

Composed: *c.*1370, N (Pic. traits)

Manuscript: BN fr. 777 [mid C15]

Edition: Bossuat (1931–33) (DEAF: BerinB)

Extract: §§1–16

Berte: (ChG, 12s). *Berte aus grans piés*, by Adenet le Roi

Composed: *c.*1275, N (Flanders)

Manuscript: Arsenal 3142 [end C13, C (IdeF) / CN (Dees: Oise, 85)]

Edition: Henry (1963) (DEAF: BerteH)

Extract: laisses I–XI (ll. 1–336)

Blondel: (Lyr, L). *Œuvre lyrique*, by Blondel de Nesle

Composed: 1175–1190, N (Pic.)

Manuscript: Arsenal 5198 (K) (nos. V, XIII); BN fr. 844 (M) (nos. VI, XII,

XVI, XX–XXII); Bern Bürgerbibliothek 389 (C) (XV) [MS K beg. C14, MS M 2nd half C13, MS C end C13; MS K loc. not given, MS M N (Pic.), MS C E (Lorr.)]

Edition: Lepage (1994) (DEAF: BlondNesleL)

Extract: chansons V, VI, XII, XIII, XV, XVI, XX–XXII

BouchAbev: (Fabliau, 8s). *Le Bouchier d'Abeville*

Composed: mid C13, N (Pic.)

Manuscript: Berlin, Deutsche Staatsbibl., Hamilton 257 (C) [*c.*1300, W (Normandy)]

Edition: Noomen and van den Boogaard (1983–98) (DEAF: BouchAbevN)

Extract: Le Bouchier d'Abeville (in vol. 3: 237–336), full text (546 lines)

Brendan: (VieS, 8s). *Le Voyage de saint Brendan*, by Benedeit

Composed: 1106–1121, Eng

Manuscript: BL Cotton Vespasian B. X (I) [2nd half C13, Eng]

Edition: Short and Merrilees (1979) (DEAF: BrendanS)

Extract: ll. 1–500

BrevNobles: (LyrNarr, 10s). *Le Breviaire des nobles*, by Alain Chartier

Composed: before 1415

Manuscript: Manchester, Chetham's Lib., Muniment A.6.91 (Oj) [mid C15]

Edition: Laidlaw (1974)

Extract: Le Breviaire des Nobles (pp. 393–410), all octosyllabic ballads

ChAntioche: (ChG, 12s). *La Chanson d'Antioche*

Composed: 1171–1181, N (Flanders)

Manuscript: BN fr. 12558 (A) [mid C13, N (Arras)]

Edition: Duparc-Quioc (1976) (DEAF: AntiocheD)

Extract: laisses I–XVI (ll. 1–320)

CharlesOrl: (Lyr, L). *Poésies*, by Charles d'Orléans

Composed: 1410–1460, S (Orleans)

Manuscript: BN fr. 25458 [*c.*1455 (ed.) / C15 (DEAF)]

Edition: Champion (1923–24) (DEAF: CharlD'OrlC)

Extract: pieces I, V, XXIII, XXV, XXIX, LXI, LXXV, LXXXI, LXXXIX, C; chansons I–III, XI–XV, XXVI, XXVIII, CCXXXIV–CCXXXVIII, CCCXLVI, CCCXLVII

CharNimes: (ChG, 10s). *Le Charroi de Nîmes*

Composed: 2nd third C12, unk.

Manuscript: BN fr. 1449 (A2) [3rd quart. C13, C (IdeF) / C (Dees: Haute-Marne, 70)]

Edition: McMillan (1978) (DEAF: CharroiM)

Extract: laisses I–XV (ll. 1–402)

Charrete: (Rom, 8s). *Le Chevalier de la charrete*, by Chrétien de Troyes

Composed: *c.*1177, C (Ch.)

Manuscript: BN fr. 794 (C) [*c.*1235, C (Ch.) / C (Dees: Haute-Marne, 79)]

Edition: Roques (1958) (DEAF: LancR)

Extract: ll. 503–1003

Chartier: (Lyr, L). *Rondeaulx et ballades*, by Alain Chartier

Composed: 1410–1425

Manuscript: Toulouse Bibl. mun. 826 (Qd) (nos. I–XXII, XXIV–XXVI); Grenoble Bibl. mun. 874 (Nj) (no. XXVII); Berlin Kupferstichkabinett, 78 B 17 (Td) (nos. XXIII, XXVIII) [MS Qd early C15, MS Nj mid C15, MS Td late C15]

Edition: Laidlaw (1974)

Extract: Rondeaulx and ballades (pp. 371–93), all

Chivalier: (Fabliau, 8s). *Un chivalier et sa dame et un clerk*

Composed: 2nd third C13, Eng

Manuscript: Corpus Christi Coll., Cambridge, 50 (h) [last quart. C13, Eng]

Edition: Noomen and van den Boogaard (1983–98) (DEAF: ChevDameClercN)

Extract: Un Chivalier et sa dame et un clerk (in vol. 10: 115–127), ll. 1–504

ChrFroiss: (Chr, P). *Chroniques*, by Jean Froissart

Composed: 1390

Manuscript: Amiens Bibl. mun. 486 [date not given]

Edition: Diller (1991–) (DEAF: FroissChronL; FroissChronIII1D)

Extract: §§2–10

Christine: (Lyr, L). *Ballades, Rondeaux and Virelais*, by Christine de Pisan

Composed: 1390–1410

Manuscript: BL Harley 4431 (A2) [*c.*1413, C (IdeF)]

Edition: Varty (1965)

Extract: balades 9, 17, 20, 21, 25, 26, 50, 53, 54, 76, 85, 102; rondeaux 30, 35, 36, 38, 55, 56, 60, 61, 65, 80, 81, 84, 115

ChRoland: (ChG, 10s). *La Chanson de Roland*

Composed: *c.*1100, Eng / W (?)

Manuscript: Oxford Bodl. Digby 23 [2nd quart. C12, Eng]

Edition: Moignet (1969) (DEAF: RolMoign)

Extract: laisses 80–110 (ll. 1017–1419)

ChvBarisel: (conte pieux, 8s). *Le Chevalier au barisel*

Composed: beg. C13, N (Pic.)

Manuscript: BN fr. 1109 [1310, N (Pic.) / N (Dees: Somme, Pas-de-Calais, 95)]

Edition: Lecoy (1955) (DEAF: ChevBarAnL)
Extract: ll. 501–1084

ComteAnjou: (Rom, 8s). *Le Roman du comte d'Anjou*, by Jehan Maillart
Composed: 1316, CN (Sth. Pic.) / C (IdeFr)
Manuscript: BN nfr. 4531 (A) [1316 or slightly later, N (Pic.)]
Edition: Roques (1931) (DEAF: MaillartR)
Extract: ll. 1–509

ComtePoit: (Rom, 8s). *Le Roman du comte de Poitiers*
Composed: 1st half C13, N (Pic.)
Manuscript: Arsenal 3527 [beg. C14, N (Pic.) / N (Dees: Somme, Pas-de-Calais, 86)]
Edition: Malmberg (1940) (DEAF: ComtePoitM)
Extract: ll. 1–500

CononBeth: (Lyr, L). *Chansons*, by Conon de Béthune
Composed: 1180–1192, N (Arras)
Manuscript: comparative edition, orth. from BN fr. 12615 (T) [1st part last quart. C13, N (Artois)]
Edition: Wallensköld (1921) (DEAF: ConBethW2)
Extract: chansons I–III, V, VII–X

ConqVilleh: (Chr, P). *La Conquête de Constantinople*, by Geoffroi de Villehardouin
Composed: before 1209, unk. (author from East Ch.)
Manuscript: Oxford Bodl. Laud Misc. 587 (O) [2nd half C14, N (Pic.) (DEAF) / E (Dees: Burgundy, 74)]
Edition: Faral (1961) (DEAF: VillehF)
Extract: ¶¶12–48

Dolopathos: (Rom, 8s). *Le Roman de Dolopathos*, by Herbert
Composed: 1223, E (Lorr.)
Manuscript: Montpellier Ec. de Méd. 436 [3rd quart. C13, E (Lorr.)]
Edition: Leclanche (1997) (DEAF: DolopL)
Extract: ll. 1–500

Eracle: (Rom, 8s). *Eracle*, by Gautier d'Arras
Composed: *c.*1175, N (Pic.)
Manuscript: BN fr. 1444 [end C13, CN (Sth. Pic.)]
Edition: Raynaud de Lage (1976) (DEAF: GautArrErH)
Extract: ll. 5519–6020

EurialLucr: (Rom, 10s). *Ystoire de Eurialus et Lucresse*, by Octovien de Saint-Gelais
Composed: *c.*1490
Manuscript: Hain-Copinger 243/244 (Paris, 1492/1493) [early printings; six

extant exemplars] [1493]
Edition: Richter (1914)
Extract: ll. 1–402

FEUILLEE: (SecPlay, 8s). *Le Jeu de la feuillee*, by Adam de la Halle
Composed: 1276, N (Pic.)
Manuscript: BN fr. 25566 (P) [prob. 1295, N (Arras) / N (Dees: Somme, Pas-de-Calais, 89)]
Edition: Langlois (1951) (DEAF: AdHaleFeuillD)
Extract: ll. 13–535

FLORIMONT: (Rom, 8s). *Florimont*, by Aimon de Varennes
Composed: 1188, SE (Fr.-Pr. traits)
Manuscript: BN fr. 15101 (F) [beg. C13, unk.]
Edition: Hilka (1932) (DEAF: AimonFlH)
Extract: ll. 965–1456

FROISSART: (Lyr, L). *Lyric Poems*, by Jean Froissart
Composed: 1362–1392, N (Pic.)
Manuscript: BN fr. 831 (A) [1394]
Edition: McGregor (1975) (DEAF: FroissPoésM)
Extract: balades V–VIII, XV, XXI, XXV, XXVI, XXVIII, XXIX, XXXIV, XXXIX; virelais II, IV, VI, IX, XII, XIII

GACEBRULE: (Lyr, L). *Chansons*, by Gace Brulé
Composed: 1180–1210, C (Sth. Ch.)
Manuscript: comparative, orth. from BN fr. 844 (M) except BN fr. 846 (O) (nos. LVI–LVIII) [MS M 2nd half C13, MS O 2nd half C13; MS M N (Pic.), MS O loc. not given]
Edition: Dyggve (1951) (DEAF: GaceBruléD)
Extract: chansons VII, X, XVI, XIX, XLIII, XXXII, XXXIV, XLV, LVI–LVIII

GORMISEM: (ChG, 8s). *Gormont et Isembart*
Composed: 1st half C12, C (IdeF)
Manuscript: Brussels Bibl. Roy. Portefeuille II 181 [1st quart. C13, Eng (Dees: 80)]
Edition: Bayot (1931), diplomatic transcription (DEAF: GormB)
Extract: ll. 8–508

GRISELDIS: (RelPlay, 8s). *L'Estoire de Griseldis*
Composed: 1395
Manuscript: BN fr. 2203 (Cangé) [date not given, N (Pic.)]
Edition: Roques (1957) (DEAF: GriseldisEstR)
Extract: ll. 1–501

GuiWarewic: (Rom, 8s). *Gui de Warewic*

Composed: 1st third C13 (DEAF) / 1232–1242 (ed.), Eng

Manuscript: Brit. Mus. Add. 38662 [2nd quart. C13, Eng]

Edition: Ewert (1932–33) (DEAF: GuiWarE)

Extract: ll. 1–500

HugCapet: (ChG, 12s). *Hugues Capet*

Composed: *c.*1360, prob. N

Manuscript: Arsenal 3145 [2nd half C15, NE (Walloon)]

Edition: Laborderie (1997) (DEAF: HugCapLb)

Extract: laisses I–VIII (ll. 1–370)

HuonBord: (ChG, 10s). *Huon de Bordeaux*

Composed: after 1216 (mid C13), N

Manuscript: Tours 936 (M) [mid C13, N]

Edition: Ruelle (1960) (DEAF: HuonR)

Extract: laisses I–IV (part) (ll. 1–402)

ImageMonde: (Didactic, 8s). *L'Image du Monde*

Composed: *c.*1248, E (Lorr.)

Manuscript: BL Harley 4333 [2nd half C13, E (Lorr.)]

Edition: Meyer (1892) (DEAF: ImMondeOct3M)

Extract: ll. 1–502

IsopetLyon: (Fables, 8s). *L'Isopet de Lyon*

Composed: 2nd half C13, E (Fr.-Com.)

Manuscript: Lyon Bibl. mun. Palais des Arts 57 (L') [end C13, E (Fr.-Com.) (Dees: 95)]

Edition: Bastin and Ruelle (1929–) (DEAF: YsLyonB)

Extract: Isopet de Lyon (in vol. 2: 83–198), prologue and fables I–VIII

JehSaintre: (Rom, P). *Jehan de Saintré*, by Antoine de la Sale

Composed: 1451–1456

Manuscript: BN nfr. 10057 (Barrois) [1456]

Edition: Eusebi (1993–94) (DEAF: JSaintréE)

Extract: chs II–III

JeuAdam: (RelPlay, 8s). *Le Jeu d'Adam (Ordo Representacionis Ade)*

Composed: 2nd half C12 (DEAF) / 1146–1174 (ed.), Eng

Manuscript: Tours 927 [2nd quart. C13, S (Tours)]

Edition: Aebischer (1963) (DEAF: AdamN)

Extract: ll. 1–48, 117–518, 623–674

JeuNicolas: (RelPlay, 8s). *Le Jeu de saint Nicolas*, by Jean Bodel

Composed: 1191–1202, N (Arras)

Manuscript: BN fr. 25566 [prob. 1295, N (Arras) / N (Dees: Somme, Pas-de-Calais, 94)]
Edition: Henry (1965) (DEAF: BodelNicH2)
Extract: ll. 1–238, 251–383, 412–540

JudithHolo: (RelPlay, 8s). *Le Mystère de Judith et Holofernes*, by Jean Molinet (attribution doubtful)
Composed: 1480–1490
Manuscript: BN rés Yf 11 (Paris: Pierre Le Dru) [*c.*1500]
Edition: Runnalls (1995)
Extract: ll. 25–519

LaisMarie: (Lais, 8s). *Lais*, by Marie de France
Composed: *c.*1165, Eng (author from IdeF)
Manuscript: BL Harley 978 [2nd half C13, Eng (Pic. traits) / Eng (75, Dees)]
Edition: Ewert (1965) (DEAF: MarieLaisE)
Extract: Lanval: ll. 1–500

LibFortune: (poème moral, 8s). *Liber Fortunæ*
Composed: 1346, unk., E traits
Manuscript: Clermont-Ferrand 356 (C) [mid C15]
Edition: Grigsby (1967) (DEAF: LFortunaeG)
Extract: ll. 1–505

LyonCor: (Treatise, 10s). *Le Lyon coronné*
Composed: 1467, E (Burgundy)
Manuscript: Madrid, Escorial L.II.23 (E) [C15]
Edition: Urwin (1958)
Extract: sections II, IV, VI

Machaut: (Lyr, L). *La Louange des dames*, by Guillaume de Machaut
Composed: 2nd third C14, C (Ch.)
Manuscript: BN fr. 22545–6 (K, Ms des Carmes) [4th quart. C14]
Edition: Chichmaref (1909) (DEAF: GuillMachC)
Extract: La Louange des dames, pieces VI, XV, XVI, XVIII, XXIII–XXVI, LII, LIII, LXII, LXXXVI, XC, XCI, CI, CVI, CLXV, CLXVI, CCXXII

MemCommyn: (Chr, P). *Mémoires*, by Philippe des Commynes
Composed: 1489–1498
Manuscript: BN nfr. 20960 (P) [beg. C16]
Edition: Calmette (1924–25) (DEAF: CommC)
Extract: book I, chs 5, 6

MirLouis: (RelNarr, P). *Les Miracles de saint Louis*, by Guillaume de Saint-Pathus
Composed: 1297, C (Seine-et-Marne, St. Pathus)
Manuscript: BN fr. 4976 [1st quart. C14, W (Dees: Eure, 86)]

Edition: Fay (1931) (DEAF: SLouisPathMirF)

Extract: miracles 5, 6, 7

MirNDCoin: (RelNarr, 8s). *Les Miracles de Nostre Dame*, by Gautier de Coinci

Composed: *c.*1227, C (Soissons)

Manuscript: BN fr. 22928 [beg. C14, CN (Sth. Pic.) / C (Dees: Marne, 79)]

Edition: Koenig (1955–70) (DEAF: CoincyI1K)

Extract: II Mir. 22 ll. 1–246; II Mir. 23 ll. 1–252

MirNDPers: (RelPlay, 8s). *Miracles de Nostre Dame par personnages*

Composed: 1345, C (Paris)

Manuscript: BN fr. 819-820 [end C14]

Edition: Paris and Robert (1876–93) (DEAF: MirNDPers7P)

Extract: Miracle de la nonne (in vol. 1: 311–40), ll. 1–621 (incl. some 4-syll lines)

MirTheoph: (RelPlay, 8s). *Le Miracle de Théophile*, by Rutebeuf

Composed: *c.*1261, C (Sth. Ch.)

Manuscript: BN fr. 837 [last quart. C13, C (IdeFr) / C (Dees: Seine-et-Marne, 85)]

Edition: Frank (1925) (DEAF: RutebTheophF)

Extract: ll. 1–159, 169–383, 540–663

Molinet: (Lyr, L). *Les Faictz et dictz*, by Jean Molinet

Composed: 1464–1507

Manuscript: Tournai 105 (A) [date not given]

Edition: Dupire (1936–39)

Extract: poésies religeuses XXII ll. 104–174, XXIV; poésies familières XVIII, XLVI, LI, LXI, LXVIII, LXIX, LXX, LXXI, LXXII, LXXIV

MutFortune: (Rom, 8s). *Le Livre de la mutacion de Fortune*, by Christine de Pisan

Composed: 1403

Manuscript: Brussels Bibl. Roy. 9508 (B) [1403]

Edition: Solente (1959–66) (DEAF: ChrPisMutS)

Extract: chs II–VIII (ll. 51–566)

OrlogeAmor: (Dit, 10s). *L'Orloge amoureus*, by Jean Froissart

Composed: 1368, N (Pic.)

Manuscript: BN fr. 830 (B) [1393]

Edition: Dembowski (1986) (DEAF: FroissOrlD)

Extract: ll. 1–424

PassClerm: (RelNarr, 8s). *La Passion de Clermont*

Composed: *c.*1000, Clermont (?) / C (IdeFr) (?) / SW (Poitou) (?)

Manuscript: Clermont-Ferrand 240 [*c.*1000, prob. Clermont]

Edition: Avalle (1962) (DEAF: PassionA)

Extract: full text (516 lines)

PassGreban: (RelPlay, 8s). *Le Mystère de la Passion*, by Arnoul Gréban

Composed: *c.*1450, W (Le Mans)

Manuscript: BN fr. 815 (B) [1458]

Edition: Jodogne (1965–83) (DEAF: GrebanJ)

Extract: ll. 1–154, 175–539

PassJongl: (RelNarr, 8s). *La Passion des Jongleurs*, by Geufroi de Paris

Composed: *c.*1243 (orig. of text end C12), C (IdeF)

Manuscript: BN fr. 1526 [2nd half C13, C (IdeF) / C (Dees: Paris, 82)]

Edition: Perry (1981) (DEAF: PassJonglGP)

Extract: ll. 1–540

PassPalat: (RelPlay, 8s). *La Passion de Palatinus*

Composed: beg. C14 (DEAF) / 1300–1315 (ed.), E (Burgundy, traits)

Manuscript: Vat. Pal. lat. 1969 [beg. C14, C (IdeFr)]

Edition: Frank (1922) (DEAF: PassPalF)

Extract: ll. 99–599

Pathelin: (SecPlay, 8s). *La Farce de maître Pierre Pathelin*

Composed: 1464–1469

Manuscript: BN rés Yf 417 (Lyon: Guillaume Le Roy) [*c.*1485]

Edition: Dufournet (1986)

Extract: ll. 1–505

Prothes: (Rom, 8s). *Protheselaus*, by Hue de Rotelande

Composed: *c.*1185, Eng, W traits (standard W French, ed.)

Manuscript: BN fr. 2169 (A) [C13, Eng]

Edition: Holden (1991–93) (DEAF: ProtH)

Extract: ll. 1–504

QuadrInvec: (Treatise, P). *Le Quadrilogue invectif*, by Alain Chartier

Composed: 1422

Manuscript: BN fr. 126 [early C15]

Edition: Droz (1950)

Extract: p. 5, l. 15 – p. 15, l. 12

QuatrLivre: (Biblical trans., P). *Li Quatre livre des reis*

Composed: 2nd half C12, Eng

Manuscript: Maz. 54 (70) [end C12, Eng]

Edition: Curtius (1911) (DEAF: RoisC)

Extract: book I, 1:1–3:21

RaoulCamI: (ChG, 10s). *Raoul de Cambrai (first part)*

Composed: end C12, NE

Manuscript: BN fr. 2493 [1st half C13, N (Pic.) (DEAF) / NE (Dees: Ardennes, 75)]
Edition: Kay (1992) (DEAF: RCambrK)
Extract: laisses XXIII–XLIII (ll. 296–745)

RenartNouv: (Rom, 8s). *Renart le nouvel*, by Jacquemars Giélée
Composed: *c.*1290, N (Arras)
Manuscript: BN fr. 25566 [prob. 1295, N (Arras) / N (Dees: Somme, Pas-de-Calais, 94)]
Edition: Roussel (1961) (DEAF: RenNouvR)
Extract: ll. 1–506

RoseMeun: (Rom, 8s). *Le Roman de la rose (second part)*, by Jean de Meun
Composed: *c.*1275, C (IdeF), S traits
Manuscript: BN fr. 1573 (Ha) [*c.*1285, SW (Orleans) / C (Dees: Paris, 89)]
Edition: Lecoy (1965–75) (DEAF: RoseMLec)
Extract: ll. 4039–4532 (beg. of Meun's continuation)

Rou: (Chr, 12s). *Le Roman de Rou*, by Wace
Composed: *c.*1170 (1160–1174), W
Manuscript: BN Duchesne 79 [C17 (copy of lost original *c.*1300), C (IdeF), W traits]
Edition: Holden (1970) (DEAF: RouH)
Extract: 2e partie, ll. 1–329

Rutebeuf: (Lyr, L). *Œuvres*, by Rutebeuf
Composed: 1249–1277, C (Sth. Ch.)
Manuscript: BN fr. 837 (A) (nos. 2, 40, 41, 51); BN fr. 1635 (C) (nos. 13, 21, 38, 39) [MS A last quart. C13, MS C end C13; MS A C (IdeFr), MS C E (East Ch. / Burgundy)]
Edition: Faral and Bastin (1959–60) (DEAF: RutebF)
Extract: pieces 2, 13, 21, 38–41, 51

Sacrist3: (Fabliau, 8s). *Le Sacristain (troisième rédaction)*
Composed: [2nd half C13], E (Burgundy)
Manuscript: BN fr. 1593 (E) [end C13, C / E (IdeFr, weak Lorr. traits)]
Edition: Noomen and van den Boogaard (1983–98) (DEAF: NoomenFabl)
Extract: Le Sacristain, version III (in vol. 7: 1–190), ll. 1–501

Testament: (satirical poem, 8s). *Le Testament*, by François Villon
Composed: 1461–1462
Manuscript: BN fr. 20041 (C) [dating impossible (ed.)]
Edition: Rychner and Henry (1974)
Extract: quatrains XCVI–CLVIII (ll. 998–1237, 1266–1377, 1406–1421, 1457–1472, 1507–1514, 1543–1590, 1628–1691; lyric pieces excluded)

Thebes: (Rom, 8s). *Le Roman de Thèbes*

Composed: *c.*1160, SW (Poitiers)

Manuscript: BN fr. 784 [2nd third C13, C (IdeF) (Dees: Paris, 81)]

Edition: Raynaud de Lage (1966) (DEAF: ThebesR)

Extract: ll. 103–606

ThibautCh: (Lyr, L). *Chansons*, by Thibaut de Champagne

Composed: 2nd quart. C13, C (Ch.)

Manuscript: Arsenal 5198 (K) except BN fr. 846 (O) (no. XXXVII), BN fr. 24406 (V) (no. LXI) [MS K beg. C14, MS O 2nd half C13, MS V last third C13; loc. not given]

Edition: Wallensköld (1925) (DEAF: ThibChampW)

Extract: chansons I, IV, X, XXIV, XXXIV, XXXVII, XXXIX, XLII, LXI

TristanPr: (Rom, P). *Le Roman de Tristan en prose*

Composed: *c.*1230, unk.

Manuscript: Vienna BN 2542 [*c.*1300, N (Pic.)]

Edition: Ménard (1987–) (DEAF: TristPrMé)

Extract: vol. 1, §§1–10

TroysGalans: (SecPlay, 8s). *Troys galans, le Monde qu'on faict paistre et Ordre*

Composed: *c.*1445

Manuscript: BN fr. 24341 (La Vallière) [*c.*1575, W (Rouen)]

Edition: Picot (1902–12) (DEAF: SottiesP)

Extract: Farce joyeuse à cinq personnages (in vol. I: 11–46), full text

VieEdmund: (VieS, 8s). *La Passiun de seint Edmund*

Composed: beg. C13 / *c.*1200 (ed.), Eng

Manuscript: Gonville and Caius Coll., Cambr., MS 435/435 [1st half C13, Eng]

Edition: Grant (1978) (DEAF: SEdmPassG)

Extract: quatrains 1–125 (ll. 1–500)

VieLeger: (VieS, 8s). *La Vie de saint Léger*

Composed: *c.*1000, SW (Poitou) (?) / Walloon reworked in Clermont (?)

Manuscript: Clermont-Ferrand 240 [*c.*1000, prob. Clermont]

Edition: Linskill (1937) (DEAF: SLégerA)

Extract: full text (240 lines)

VieLouis: (Chr, P). *La Vie de saint Louis*, by Jehan de Joinville

Composed: 1305–1309, C (Ch.)

Manuscript: BN fr. 13568 (A) [*c.*1335, NE (DEAF) / C (Paris, 87, Dees)]

Edition: Corbett (1977) (DEAF: JoinvC)

Extract: ¶¶19–45

VieRichard: (VieS, 8s). *La Vie seint Richard evesque de Cycestre*, by Pierre D'Abernon

of Fetcham

Composed: *c.*1270, Eng

Manuscript: BL Add. 70513 [1st quart. C14, Eng]

Edition: Russell (1995) (DEAF: PAbernRichR)

Extract: ll. 1–516

Villon: (Lyr, L). *(Lyric verse)*, by François Villon

Composed: 1457–1463

Manuscript: various

Edition: Rychner and Henry (1977) (DEAF: VillonLaisR)

Extract: poèmes variés (in vol. 1: 33–77), nos. 1, 2, 4–7, 10–13

VoirDit: (Dit, 8s). *Le Livre du voir dit*, by Guillaume de Machaut

Composed: 1364, C (Ch.)

Manuscript: BN fr. 22545 (Ms des Carmes) [4th quart. C14]

Edition: Imbs (1999) (DEAF: GMachVoirI)

Extract: ll. 1–202, 216–373, 387–474, 488–535 (lyrics excluded)

References

Abeillé, Anne, Lionel Clément, and François Toussenel. 2003. 'Building a Treebank for French', in Anne Abeillé (ed.) *Treebanks: Building and Using Parsed Corpora*, Text, Speech and Language Technology 20 (Dordrecht: Kluwer), pp. 165–88

Adams, Marianne. 1987a. 'From Old French to the Theory of Pro-drop', *Natural Language and Linguistic Theory*, 5: 1–32

——. 1987b. 'Old French, Null Subjects, and Verb Second Phenomena', unpublished doctoral dissertation, University of California, Los Angeles

——. 1989. 'Verb Second Effects in Medieval French', in Carl Kirschner and Janet DeCesaris (eds) *Studies in Romance Linguistics: Selected Proceedings from the Seventeenth Linguistic Symposium on Romance Languages (XVII. LSRL), Rutgers University, 27–29 March 1987*, Current Issues in Linguistic Theory 60 (Amsterdam; Philadelphia: Benjamins), pp. 1–31

Aebischer, Paul (ed.). 1963. *Le Mystère d'Adam (Ordo Representacionis Ade): Texte complet du manuscrit de Tours*, Textes littéraires français 99 (Geneva: Droz)

Altmann, Barbara K. (ed.). 1998. *The Love Debate Poems of Christine de Pizan* (Gainesville: University Press of Florida)

Armstrong, Edward Cooke (ed.). 1937. *The Medieval French Roman d'Alexandre*, 7 vols (Princeton: Princeton University Press)

Avalle, D'Arco Silvio. 1962. *Cultura e lingua francese delle origini nella 'Passion'* (Milan: Ricciardi)

Ayres-Bennett, Wendy. 1996. *A History of the French Language Through Texts* (London: Routledge)

Banniard, Michel. 1998. 'Le Problème de l'accent en français: Remarques sur son origine et sur son évolution', in J. Claude Faucon and Danielle Quéruel (eds) *Miscellanea Mediaevalia: Mélanges offerts à Philippe Ménard*, 2 vols, Nouvelle Bibliothèque du Moyen Age 46 (Paris: Champion), vol. 1, pp. 65–73

Bastin, Julia, and Pierre Ruelle (eds). 1929–. *Recueil général des Isopets*, 4 vols, Société des anciens textes français (Paris: Champion)

Batllori, Montserrat, and Francesc Roca (eds). 2005. *Grammaticalization and Parametric Variation* (Oxford; New York: Oxford University Press)

Battye, Adrian, and Ian Roberts (eds). 1995. *Clause Structure and Language Change*, Oxford Studies in Comparative Syntax (New York; Oxford: Oxford University Press)

Bayot, Alphonse (ed.). 1931. *Gormont et Isembart: Fragment de chanson de geste du XIIe siècle*, 3rd rev. edn, Classiques français du Moyen Age 14 (Paris: Champion)

Beaulieux, Charles. 1927. *Histoire de l'orthographe française*, 2 vols (Paris: Champion)

Beaussart, François-Jérôme. 1989. 'Les Miracles de Gautier de Coinci', in F.-J. Beaussart and J. Subrenat (eds) *Miracles et mystères: La Littérature religieuse au nord de la France*, Trésors littéraires du nord de la France 6 (Troesnes: Corps 9), pp. 13–108

Beckman, Mary E. 1986. *Stress and Non-Stress Accent*, Netherlands Phonetics Archives 7 (Dordrecht; Riverton, NJ: Foris)

Bédier, Joseph. 1908–13. *Les Légendes épiques*, 4 vols (Paris: Champion)

——. 1925. *Les Fabliaux : Etudes de littérature populaire et d'histoire littéraire du moyen-âge*, 5th edn (Paris: Champion)

—— (ed. and trans.). 1928. *La Chanson de Roland publiée d'après le manuscrit d'Oxford*, 66th edn (Paris: L'Edition d'art, H. Piazza)

Benincà, Paola. 1995. 'Complement Clitics in Medieval Romance: The Tobler-Mussafia Law', in Battye and Roberts (1995), pp. 325–44

——. 2001. 'The Position of Topic and Focus in the Left Periphery', in Guglielmo Cinque and Giampaolo Salvi (eds) *Current Studies in Italian Syntax: Essays Offered to Lorenzo Renzi*, North Holland Linguistic Series 59 (Amsterdam; London: Elsevier), pp. 39–64

——. 2006. 'A Detailed Map of the Left Periphery of Medieval Romance', in Raffaella Zanuttini (ed.) *Crosslinguistic Research in Syntax and Semantics: Negation, Tense, and Clausal Architecture*, Georgetown University Round Table on Languages and Linguistics (2004) (Washington, D.C.: Georgetown University Press), pp. 53–86

Benincà, Paola, and Cecilia Poletto. 2004. 'Topic, Focus and V2: Defining the CP Sublayers', in Rizzi (2004), pp. 52–75

Benveniste, Emile. 1966. *Problèmes de linguistique générale*, 2 vols (Paris: Gallimard)

Berthelot, Anne. 2006. *Histoire de la littérature française du moyen âge*, Collection Histoire de la littérature française (Rennes: Presses Universitaires de Rennes)

BFM. 2005. *Base de français médiéval*, (Lyons: UMR ICAR / ENS-LSH) <`http://bfm.ens-lsh.fr`> [accessed 20 January 2011]

Bies, Ann, Mark Ferguson, Karen Katz, and Robert MacIntyre. 1995. 'Bracketing Guidelines for Treebank II Style: Penn Treebank Project', technical manual, University of Pennsylvania <`ftp://ftp.cis.upenn.edu/pub/treebank/doc/manual/root.ps.gz`> [accessed 20 January 2011]

Billy, Dominique. 1999. 'Le Décasyllabe rénové de Froissart', *Romania*, 117: 508–44

Blanche-Benveniste, Claire. 1996. 'Trois remarques sur l'ordre des mots dans la langue parlée', *Langue Française*, 111: 109–24

Blum, Claude (ed.). 2001. *Corpus de la littérature médiévale en langue d'oïl des origines à la fin du XVe siècle* (Paris: Champion électronique) [on CD]. Online version <`http://www.classiques-garnier.com/numerique/index.php`> [accessed 20 January 2011]

Bossuat, Robert (ed.). 1931–33. *Bérinus: Roman en prose du XIVe siècle*, 2 vols (Paris: Société des anciens textes français)

Bruckner, Matilda Tomaryn. 2000. 'The Shape of Romance in Medieval France', in Roberta L. Krueger (ed.) *The Cambridge Companion to Medieval Romance*, Cambridge Companions to Culture (Cambridge: Cambridge University Press), pp. 13–28

Buridant, Claude. 1987. 'Les Résidus de l'ordre OV en ancien français et leur effacement en moyen français', *Romania*, 108: 20–65

——. 1992. 'L'Evolution de l'ancien français vers le français contemporain', in Gérold Hilty (ed.) *Actes du XXe Congrès international de linguistique et philologie romanes: Université de Zurich (6–11 avril 1992)*, 5 vols (Tübingen: Francke), vol. 3, pp. 25–49

——. 2000. *Grammaire nouvelle de l'ancien français* (Paris: Sedes)

Besten, Hans den. 1983. 'On the Interaction of Root Transformations and Lexical Deletive Rules', in Werner Abrams (ed.) *On the Formal Syntax of the Westgermania: Papers from the '3rd Groningen Grammar Talks,' Groningen, January 1981* (Amsterdam: Benjamins), pp. 47–131

Cailly, Marie. 2007. *Les Fabliaux, la satire et son public : L'Oralité dans la poésie satirique et profane en France, XIIe–XIVe siècles* (Cahors: La Louve)

Calin, William C. 1961. 'Sur la composition de Gormond et Isembard', *Cultura Neolatina*, 21: 57–65

Calmette, Joseph (ed.). 1924–25. *Mémoires*, by Philippe de Commynes, 3 vols, Classiques de l'histoire de France au Moyen Age 3, 5–6 (Paris: Champion)

Cardinaletti, Anna, and Ian Roberts. 1991. 'Clause Structure and X-Second', manuscript, University of Geneva. Published in Guglielmo Cinque (ed.). 2002. *Functional Structure in DP and IP: The Cartography of Syntactic Structures, vol. 1*, Oxford Studies in Comparative Syntax (Oxford: Oxford University Press), pp. 123–66

Castellani, Arrigo. 1969. 'L'Ancien Poitevin et le problème linguistique des Serments de Strasbourg', *Cultura Neolatina*, 29: 201–34

——. 1978. 'Nouvelles Remarques au sujet de la langue des Serments de Strasbourg', in André Gendre, Charles-Théodore Gossen, and Georges Straka (eds) *Mélanges d'études romanes du Moyen Age et de la Renaissance offerts à Monsieur Jean Rychner*, Travaux de linguistique et de littérature 26(1) (Strasbourg: Centre de Philologie et Littératures Romanes de l'Université de Strasbourg), pp. 61–73

Cerquiglini, Bernard. 1989. *Eloge de la variante: Histoire critique de la philologie* (Paris: Seuil)

Champion, Pierre (ed.). 1923–24. *Poésies*, by Charles d'Orléans, 2 vols, Classiques français du Moyen Age 34, 56 (Paris: Champion)

Chichmaref, Vladimir (ed.). 1909. *Poésies lyriques*, by Guillaume de Machaut, 2 vols (Paris: Champion)

Combettes, Bernard. 1988. 'Recherches sur l'ordre des éléments de la phrase en moyen français', unpublished habilitation thesis, Université de Nancy

Connell, Bruce and Amalia Arvaniti (eds). 1995. *Phonology and Phonetic Evidence*, Papers in Laboratory Phonology 4 (Cambridge: Cambridge University Press)

Corbett, Noël L. (ed.). 1977. *La Vie de saint Louis: Le Témoignage de Jehan, seigneur de Joinville* (Sherbrooke: Naaman)

Curtius, Ernst Robert (ed.). 1911. *Li Quatre Livre des Reis: Die Bücher Samuelis und der Könige in einer französischen Bearbeitung des 12. Jahrhunderts, nach der ältesten Handschrift unter Benutzung der neu aufgefundenen Handschriften*,

Gesellschaft für romanische Literatur 26 (Dresden: Gedruckt für die Gesellschaft für romanische literatur)

Dauer, Rebecca M. 1983. 'Stress-Timing and Syllable-Timing Reanalyzed', *Journal of Phonetics*, 11: 51–62

Dees, Anthonij. 1980. *Atlas des formes et des constructions des chartes françaises du 13e siècle*, Beihefte zur Zeitschrift für romanische Philologie 178 (Tübingen: Niemeyer)

——. 1987. *Atlas des formes linguistiques des textes littéraires de l'ancien français*, Beihefte zur Zeitschrift für romanische Philologie 212 (Tübingen: Niemeyer)

Delais-Roussarie, Elisabeth. 1995. 'Pour une approche probabiliste de la structure prosodique: Etude de l'organisation prosodique et rythmique de la phrase française', unpublished doctoral thesis, Université de Toulouse-Le-Mirail

Dell, François. 1984. 'L'Accentuation dans les phrases en français', in François Dell, Daniel Hirst and Jean-Roger Vergnaud (eds) *Forme sonore du langage : Structure des représentations en phonologie* (Paris: Hermann), pp. 65–122

Dembowski, Peter F. (ed.). 1969. *Ami et Amile: Chanson de geste*, Classiques français du Moyen Age 97 (Paris: Champion)

—— (ed.). 1986. *Le Paradis d'amour; L'Orloge amoureus*, by Jean Froissart, Textes littéraires français 339 (Geneva: Droz)

De Poerck, Guy. 1955. 'Le Sermon bilingue sur Jonas du ms. Valenciennes 521 (475)', *Romanica Gandensia*, 4: 31–66

——. 1963. 'Les Plus Anciens Textes de la langue française comme témoins de l'époque', *Revue de linguistique romane*, 27: 1–34

Di Cristo, Albert. 1999. 'Vers une modélisation de l'accentuation du français, I', *Journal of French Language Studies*, 9: 143–79

——. 2000. 'Vers une modélisation de l'accentuation du français (seconde partie)', *Journal of French Language Studies*, 10: 27–44

Di Cristo, Albert, and Daniel Hirst. 1993. 'Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français', *Travaux de l'Institut de phonétique d'Aix-en-Provence*, 15: 9–24

——. 1997. 'L'Accentuation non-emphatique en français: Stratégies et paramètres', in Jean Perrot (ed.) *Polyphonie pour Iván Fónagy* (Paris: L'Harmattan), pp. 71–101

Diller, Georges T. (ed.). 1991–. *Chroniques*, by Jean Froissart, 5 vols, Textes littéraires français 407, 415, 424, 429, 499 (Geneva: Droz)

Droz, Eugénie (ed.). 1950. *Le Quadrilogue invectif*, by Alain Chartier, 2nd rev. edn, Classiques français du Moyen Age 32 (Paris: Champion)

Duffell, Martin J. 1996. 'Chaucer, Gower and the History of the Hendecasyllable', in C. B. McCully and J. J. Anderson (eds) *English Historical Metrics* (Cambridge: Cambridge University Press), pp. 210–18

——. 2008. *A New History of English Metre*, Studies in Linguistics 5 (London: Legenda)

Dufournet, Jean (ed. and trans.). 1986. *La Farce de maître Pierre Pathelin*, Garnier Flammarion 462 (Paris: Flammarion)

Dufresne, Monique. 1993. 'L'Articulation syntaxique et phonologique de la cliticisation: Le Cas des pronoms sujets en moyen français', unpublished doctoral dissertation, University of Quebec

——. 1995. 'Etude diachronique de la cliticisation phonologique des pronoms sujets en moyen français', *Revue québécoise de linguistique*, 24: 83–109

Dufresne, Monique, and Fernande Dupuis. 1994. 'Modularity and the Reanalysis of the French Subject Pronoun', *Probus*, 6: 103–22

Duggan, Joseph J. 1973. *The Song of Roland: Formulaic Style and Poetic Craft*, Publications of the Center for Medieval and Renaissance Studies 6 (Berkeley: University of California Press)

——. 1989. 'Performance and Transmission, Aural and Ocular Reception in the Twelfth- and Thirteenth-Century Vernacular Literature of France', *Romance Philology*, 43: 49–58

Duparc-Quioc, Suzanne (ed.). 1976. *La Chanson d'Antioche*, 2 vols, Documents relatifs à l'histoire des croisades 11–12 (Paris: Librairie Orientaliste Paul Geuthner)

Dupire, Noël (ed.). 1936–39. *Les Faictz et dictz de Jean Molinet*, 3 vols, (Paris: Société des anciens textes français)

Dupoux, Emmanuel, Christophe Pallier, and Núria Sebastián. 1997. 'A Destressing "Deafness" in French?', *Journal of Memory and Language*, 36: 406–21

Dupoux, Emmanuel, and Sharon Peperkamp. 2002. 'Fossil Markers of Language Development: Phonological "Deafnesses" in Adult Speech Processing', in Jacques

Durand and Bernard Laks (eds) *Phonetics, Phonology, and Cognition*, Oxford Studies in Theoretical Linguistics (New York: Oxford University Press), vol. 3, pp. 168–90

Dupoux, Emmanuel, Sharon Peperkamp, and Núria Sebastián-Gallés. 2001. 'A Robust Method to Study Stress "Deafness"', *Journal of the Acoustical Society of America*, 110: 1606–18

Dupoux, Emmanuel, Núria Sebastián-Gallés, Eduardo Navarrete, and Sharon Peperkamp. 2008. 'Persistent Stress "Deafness": The Case of French Learners of Spanish', *Cognition*, 106: 682–706

Dupoux, Emmanuel, Núria Sebastián-Gallés, and Sharon Peperkamp. 2010. 'Limits on Bilingualism Revisited: Stress "Deafness" in Simultaneous French–Spanish Bilinguals', *Cognition*, 114: 266–75

Dupuis, Fernande. 1989. 'L'Expression du sujet dans les propositions subordonnées en ancien français', unpublished doctoral dissertation, Université de Montréal

Dyggve, Holger Peterson. 1951. *Gace Brulé, trouvère Champenois: Edition des chansons et étude historique* (Helsinki: Société Néophilologique)

Elwert, W. Theodor. 1965. *Traité de versification française des origines à nos jours*, Bibliothèque française et romane: Série A, Manuels et études linguistiques 8 (Paris: Klincksieck)

Eusebi, Mario (ed.). 1993–94. *Saintré*, by Antoine de la Sale, 2 vols, Classiques français du Moyen Age 114–15 (Paris: Champion)

Ewert, Alfred (ed.). 1932–33. *Gui de Warewic: Roman du XIIIe siècle*, 2 vols, Classiques français du Moyen Age 74–75 (Paris: Champion)

——. 1961. *The French Language*, 2nd edn (London: Faber & Faber)

—— (ed.). 1965. *Lais*, by Marie de France, Blackwell's French texts (Oxford: Blackwell)

Fabb, Nigel, and Morris Halle. 2005. *Meter in Poetry: A New Theory* (Cambridge: Cambridge University Press)

Faral, Edmond (ed. and trans.). 1961. *La Conquête de Constantinople*, by Geoffroi de Villehardouin, 2nd rev. and corr. edn, Classiques de l'histoire de France au Moyen Age 18–19 (Paris: Belles Lettres)

Faral, Edmond, and Julia Bastin (eds). 1959–60. *Œuvres complètes de Rutebeuf*, 2 vols (Paris: A. & J. Picard)

Fay, Percival B. (ed.). 1931. *Les Miracles de saint Louis*, by Guillaume de Saint-Pathus, Classiques français du Moyen Age 70 (Paris: Champion)

Ferraresi, Gisella, and Maria Goldbach. 2002. 'V2 Syntax and Topicalisation in Old French', *Linguistische Berichte*, 18: 3–25

Féry, Caroline. 2001. 'Focus and Phrasing in French', in Caroline Féry and Wolfgang Sternefeld (eds) *Audiatur vox sapientae : A Festschrift for Arnim von Stechow*, Studia Grammatica 52 (Berlin: Akademie Verlag), pp. 153–81

Fleischman, Suzanne. 1991. 'Discourse Pragmatics and the Grammar of Old French: A Functional Representation of *si* and the Personal Pronouns', *Romance Philology*, 44: 251–83

——. 1992. 'Discourse and Diachrony: The Rise and Fall of Old French *si*', in Marinel Gerritsen and Dieter Stein (eds) *Internal and External Factors in Syntactic Change* (Berlin: de Gruyter), pp. 433–73

Fónagy, Ivan. 1980. 'L'Accent français: Accent probabilitaire', in Fónagy and Léon (1980), pp. 123–233

Fónagy, Ivan and Pierre R. Léon (eds). 1980. *L'Accent en français contemporain*, Studia Phonetica 15 (Ottawa: Didier)

Fontaine, Carmen. 1985. 'Application des méthodes quantitatives en diachronie: L'Inversion du sujet en français', unpublished M.A. thesis, Université de Québec à Montréal

Fouché, Pierre. 1952–69. *Phonétique historique du français*, 3 vols, 2nd rev. and corr. edn (Paris: Klincksieck)

Foulet, Lucien. 1924. 'Accent tonique et l'ordre des mots: Formes faibles du pronom personnel après le verbe', *Romania*, 50: 54–93

——. 1930. *Petite Syntaxe de l'ancien français*, 3rd rev. edn, Classiques français du Moyen Age 21 (Paris: Champion)

Frank, Grace (ed.). 1922. *La Passion du Palatinus: Mystère du XIVe siècle*, Classiques français du Moyen Age 30 (Paris: Champion)

—— (ed.). 1925. *Le Miracle de Théophile: Miracle du XIIIe siècle*, by Rutebeuf, Classiques français du Moyen Age 49 (Paris: Champion)

Frappier, Jean (ed.). 1954. *La Mort le roi Artu: Roman du XIIIe siècle*, Textes littéraires français (Geneva: Droz)

Garde, Paul. 1968. *L'Accent* (Paris: Presses Universitaires de France)

Gasparov, Mikhail Leonovich. 1987. 'A Probability Model of Verse (English, Latin, French, Italian, Spanish, Portuguese)', *Style*, 21: 322–58

——. 1996. *A History of European Versification*, edited by G. S. Smith and Leofranc Holford-Strevens, translated by G. S. Smith and Marina Tarlinskaja (Oxford: Clarendon)

Gossen, Charles-Théodore. 1962. 'De l'histoire des langues écrites régionales du domaine d'oïl', *Revue de linguistique romane*, 27: 271–84

——. 1967. *Französische Skriptastudien* (Vienna: Österreichische Akademie der Wissenschaften)

Grabe, Esther, and Ee Ling Low. 2002. 'Durational Variability in Speech and the Rhythm Class Hypothesis', in Gussenhoven and Warner (2002), pp. 515–46

Grabe, Esther, and Paul Warren. 1995. 'Stress Shift: Do Speakers do it or do Listeners Hear it?', in Connell and Arvaniti (1995), pp. 95–110

Grammont, Maurice. 1930. *Traité pratique de prononciation française*, 7th rev. and expanded edn, (Paris: Librairie Delagrave)

Grant, Judith (ed.). 1978. *La Passiun de seint Edmund*, Anglo-Norman Texts 36 (London: Anglo-Norman Text Society)

Grigsby, John L. (ed.). 1967. *The Middle French Liber Fortunæ*, Publications in Modern Philology 81 (Berkeley: University of California Press)

Guillot, Céline, and Sophie Prévost. 2010a. 'Manuel de référence du jeu Cattex09', technical manual, UMR ICAR, CNRS / ENS-LSH <`http://bfm.ens-lyon.fr/article.php3?id_article=323`> [accessed 20 January 2011]

——. 2010b. 'Principes d'annotation Cattex09', technical manual, UMR ICAR, CNRS / ENS-LSH <`http://bfm.ens-lyon.fr/article.php3?id_article=173`> [accessed 20 January 2011]

Gussenhoven, Carlos. 1991. 'The English Rhythm Rule as an Accent Deletion Rule', *Phonology*, 8: 1–35

——. 2004. *The Phonology of Tone and Intonation* (Cambridge: Cambridge University Press)

Gussenhoven, Carlos and Natasha Warner (eds). 2002. *Laboratory Phonology, 7* (Berlin: Mouton de Gruyter)

Guthrie, Stephen R. 1987. 'Machaut and the Octosyllabe', *Studies in the Literary Imagination*, 20: 55–75

Haan, Germen de, and Fred Weerman. 1986. 'Finiteness and Verb Fronting in Frisian', in Hubert Haider and Martin Prinzhorn (eds) *Verb Second Phenomena in Germanic Languages*, Publications in Language Sciences 21 (Dordrecht: Foris), pp. 77–110

Halle, Morris, and Samuel Jay Keyser. 1971. *English Stress: Its Form, its Growth, and its Role in Verse* (New York: Harper & Row)

Hanson, Kristin, and Paul Kiparsky. 1996. 'A Parametric Theory of Poetic Meter', *Language*, 72: 287–335

Harris, Martin. 1978. *The Evolution of French Syntax: A Comparative Approach*, Longman Linguistics Library 22 (London: Longman)

Hayes, Bruce. 1984. 'The Phonology of Rhythm in English', *Linguistic Inquiry*, 15: 33–74

——. 1995. *Metrical Stress Theory: Principles and Case Studies* (Chicago: University of Chicago Press)

Henry, Albert (ed.). 1963. *Les Œuvres d'Adenet le Roi vol. 4: Berte aus grans piés*, Travaux de la Faculté de philosophie et lettres 23 (Brussels: Presses Universitaires)

—— (ed.). 1965. *Le Jeu de saint Nicolas de Jehan Bodel*, 2nd rev. edn, Travaux de la Faculté de philosophie et lettres 21 (Brussels: Presses Universitaires)

Herman, József. 1990 [1954]. 'Recherches sur l'ordre des mots dans les plus anciens textes français en prose', in *Du latin aux langues romanes: Etudes de linguistique historique* (Tübingen: Niemeyer), pp. 234–88. Originally printed in *Acta Linguistica Academiæ Scientiarum Hungaricæ*, 4: 69–94, 351–82 (1954)

Hilka, Alfons (ed.). 1932. *Florimont: Ein Altfranzösischer Abenteuerroman*, by Aimon de Varennes, Gesellschaft für Romanische Literatur 48 (Göttingen: Gedruckt für die Gesellschaft für Romanische Literatur)

Hoepffner, Ernest (ed.). 1908–21. *Œuvres de Guillaume de Machaut*, 3 vols, Société des anciens textes français (Paris: Firmin Didot)

Holden, Anthony J. (ed.). 1970. *Le Roman de Rou de Wace*, 3 vols, Société des anciens textes français (Paris: A. & J. Picard)

—— (ed.). 1991–93. *Protheselaus*, by Hue de Rotelande, 3 vols, Anglo-Norman Texts 47–49 (London: Anglo-Norman Text Society)

Holmberg, Anders. 2000. 'Scandinavian Stylistic Fronting: How Any Category Can Become an Expletive', *Linguistic Inquiry*, 31: 445–83

Horne, Merle. 1990a. 'The Clitic Group as a Prosodic Category in Old French', *Lingua*, 82: 1–14

——. 1990b. 'Empirical Evidence for a Deletion Analysis of the Rhythm Rule', *Linguistics*, 28: 959–81

Hoskins, Steven. 1994. 'Secondary Stress and Stress Clash Resolution in French: An Empirical Investigation', in Michael L. Mazzola (ed.) *Issues and Theory in Romance Linguistics: Selected Papers from the Linguistic Symposium on Romance Languages XXIII, April 1–4, 1993* (Washington, D.C.: Georgetown University Press), pp. 35–47

Huet, Gédéon (ed.). 1902. *Chansons de Gace Brulé*, Société des anciens textes français (Paris: Firmin Didot)

Imbs, Paul (ed. and trans.). 1999. *Le Livre du voir dit = Le Dit véridique*, by Guillaume de Machaut, Livre de poche 4557 (Paris: Librairie générale française)

Jackendoff, Ray. 1977. *X Syntax: A Study of Phrase Structure*, Linguistic Inquiry Monographs 2 (Cambridge, Mass.: M.I.T. Press)

Jacobs, Haike. 1992. 'The Interaction between the Evolution of Syllable Structure and Foot Structure in the Evolution from Classical Latin to Old French', in Laeufer and Morgan (1992), pp. 55–79

——. 1993. 'The Phonology of Enclisis and Proclisis in Gallo-Romance and Old French', in William J. Ashby, Marianne Mithun, Giorgio Perissinotto, and Eduardo Raposo (eds) *Linguistic Perspectives on the Romance Languages: Selected Papers from the 21st Linguistic Symposium on Romance Languages (LSRL XXI), Santa Barbara, 21–24 February 1991*, Current Issues in Linguistic Theory 103 (Amsterdam: Benjamins), pp. 149–64

Jodogne, Omer. 1965. 'Recherches sur les débuts du théâtre religieux en France', *Cahiers de civilisation médiévale*, 8: 1–24, 179–89

—— (ed.). 1965–83. *Le Mystère de la passion*, by Arnoul Gréban, 2 vols, Mémoires de la Classe des lettres, collection in-4o 2e sér. 12(3), 13(2) (Brussels: Académie royale de Belgique)

Johnston, Ronald Carlyle. 1980. 'Matthew Paris, Jordan Fantosme, and Anglo-Norman Versification', in *Mélanges de langue et littérature françaises du Moyen*

Age et de la Renaissance: Offerts à Charles Foulon, 2 vols (Rennes: Institut de français, Université de Haute-Bretagne), vol. 1, pp. 165–70

——. 1983. 'On Scanning Anglo-Norman Verse', *Anglo-Norman Studies*, 5: 153–64

Jun, Sun-Ah. 1998. 'The Accentual Phrase in the Korean Prosodic Hierarchy', *Phonology*, 15: 189–226

Jun, Sun-Ah, and Cécile Fougeron. 1995. 'The Accentual Phrase and the Prosodic Structure of French', *Proceedings International Congress of the Phonetic Sciences*, 13: 722–25

——. 2000. 'A Phonological Model of French Intonation', in Antonis Botinis (ed.) *Intonation: Analysis, Modelling and Technology* (Dordrecht: Kluwer), pp. 209–42

——. 2002. 'Realizations of Accentual Phrase in French Intonation', *Probus*, 14: 147–72

Kay, Sarah (ed. and trans.). 1992. *Raoul de Cambrai* (Oxford: Clarendon)

Kiparsky, Paul. 1977. 'The Rhythmic Structure of English Verse', *Linguistic Inquiry*, 8: 189–247

Klausenburger, Jürgen. 1970. *French Prosodics and Phonotactics: An Historical Typology*, Beihefte zur Zeitschrift für romanische Philologie 124 (Tübingen: Niemeyer)

Klavans, Judith L. 1985. 'The Independence of Syntax and Phonology in Cliticization', *Language*, 61: 95–120

Knudson, Charles A., and Jean Misrahi (eds). 1965. *Jehan de Saintré*, by Antoine de la Sale, Textes littéraires français 117 (Geneva: Droz)

Koenig, V. Frédéric (ed.). 1955–70. *Les Miracles de Nostre Dame*, by Gautier de Coinci, 4 vols, Textes littéraires français 64, 95, 131, 176 (Geneva: Droz)

Kok, Ans de. 1985. *La Place du pronom personnel régime conjoint en français: Une étude diachronique* (Amsterdam: Rodopi)

Kroch, Anthony. 1989. 'Reflexes of Grammar in Patterns of Language Change', *Language Variation and Change*, 1: 199–244

Kroch, Anthony, and Ann Taylor. 2000. *Penn-Helsinki Parsed Corpus of Middle English*, 2nd edn <`http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-3/index.html`> [accessed 20 January 2011]

Kučera, Henry, and W. Nelson Francis. 1967. *Computational Analysis of Present-day American English* (Providence, RI: Brown University Press)

Kukenheim, Louis. 1971. 'Rôle de la prosodie dans l'histoire de la langue française', in Irénée Marcel Cluzel and François Pirot (eds) *Mélanges de philologie romane dédiés à la mémoire de Jean Boutière*, 2 vols (Liège: Soledi), vol. 1, pp. 317–31

Labelle, Marie. 2007. 'Clausal Architecture in Early Old French', *Lingua*, 117: 289–316

Labelle, Marie, and Paul Hirschbühler. 2005. 'Changes in Clausal Organization and the Position of Clitics in Old French', in Batllori and Roca (2005), pp. 60–71

Laborderie, Noëlle (ed.). 1997. *Hugues Capet: Chanson de geste du XIVe siècle*, Classiques français du Moyen Age 122 (Paris: Champion)

Lacheret-Dujour, Anne, and Frédéric Beaugendre. 1999. *La Prosodie du français* (Paris: CNRS Editions)

Ladd, D. Robert. 2008. *Intonational Phonology*, 2nd edn (Cambridge: Cambridge University Press)

Laeufer, Christine and Terrell A. Morgan (eds). 1992. *Theoretical Analyses in Romance Linguistics*, Current Issues in Linguistic Theory 74 (Amsterdam: Benjamins)

Laidlaw, James Cameron (ed.). 1974. *The Poetical Works of Alain Chartier* (London: Cambridge University Press)

Langlois, Ernest (ed.). 1925. *Le Couronnement de Louis: Chanson de geste du XIIe siècle*, 2nd edn, Classiques français du Moyen Age 22 (Paris: Champion)

—— (ed.). 1951. *Le Jeu de la feuillée*, by Adam le Bossu (Adam de la Halle), 2nd rev. edn, Classiques français du Moyen Age 6 (Paris: Champion)

Le Mée, Katharine W. 1978. *A Metrical Study of Five Lais of Marie de France*, De proprietatibus litterarum 85 (Hague; New York: Mouton)

Leclanche, Jean-Luc (ed.). 1997. *Le Roman de Dolopathos*, by Herbert, 3 vols, Classiques français du Moyen Age 124–26 (Paris: Champion)

Lecoy, Félix (ed.). 1955. *Le Chevalier au barisel: Conte pieux du XIIIe siècle*, Classiques français du Moyen Age 82 (Paris: Champion)

—— (ed.). 1965–75. *Le Roman de la rose*, by Guillaume de Lorris and Jean de Meun, 3 vols, Classiques français du Moyen Age 95, 95, 98 (Paris: Champion)

Ledgeway, Adam. 2008. 'Satisfying V2 in Early Romance: Merge vs. Move', *Journal of Linguistics*, 44: 437–70

Lemieux, Monique, and Fernande Dupuis. 1995. 'The Locus of Verb Movement in Non-Asymmetric Verb Second Languages: The Case of Middle French', in Battye and Roberts (1995), pp. 80–109

Lepage, Yvan G. (ed.). 1994. *L'Œuvre lyrique de Blondel de Nesle: Textes*, Nouvelle Bibliothèque du Moyen Age 22 (Paris: Champion)

Liberman, Mark, and Alan Prince. 1977. 'On Stress and Linguistic Rhythm', *Linguistic Inquiry*, 8: 249–336

Linskill, Joseph. 1937. *Saint Léger: Etude de la langue du manuscrit de Clermont-Ferrand* (Paris: Droz)

Lodge, R. Anthony. 1993. *French: From Dialect to Standard* (London: Routledge)

Lote, Georges. 1949–96. *Histoire du vers français*, 9 vols (Paris: Boivin [1–3]; Aix-en-Provence: Université de Provence [4–9])

Maling, Joan. 1980. 'Inversion in Embedded Clauses in Modern Icelandic', *Íslenskt mál og málfræði*, 2: 175–93

Matthews, Peter Hugoe. 2007. *The Concise Oxford Dictionary of Linguistics*, 2nd edn (Oxford: Oxford University Press)

Malmberg, Bertil (ed.). 1940. *Le Roman du comte de Poitiers: Poème français du XIIIe siècle*, Etudes romanes de Lund 1 (Lund: C. W. K. Gleerup; Copenhagen: E. Munksgaard)

Marchello-Nizia, Christiane. 1985. *Dire le vrai: L'Adverbe 'si' en français médiéval: Essai de linguistique historique* (Geneva: Droz)

——. 1995. *L'Evolution du français: Ordre des mots, démonstratifs, accent tonique* (Paris: Armand Colin)

——. 2000. 'Le Tragique Destin de *moult* en français: Changement linguistique et structures sémantico-cognitives', in Annick Englebert, Michel Pierrard, Laurence Rosier, and Dan van Raemdonck (eds) *Actes du XXIIe Congrès international de linguistique et de philologie romanes, Brussels, 23–29 juillet 1998*, 9 vols (Tübingen: Niemeyer), vol. 2, pp. 285–96

——. 2006. *Grammaticalisation et changement linguistique* (Brussels: De Boeck & Larcier)

Marnette, Sophie. 1998. *Narrateur et points de vue dans la littérature française médiévale: Une approche linguistique* (Bern: Peter Lang)

Marshall, John Henry (ed.). 1971. *The Chansons of Adam de la Halle* (Manchester: Manchester University Press)

Martin, Ernest (ed.). 1882–87. *Le Roman de Renart*, 3 vols (Strasbourg: Trübner)

Martineau, France. 2009. 'Protocole d'étiquetage GTRC: Parties du discours (PDD)', technical manual, University of Ottawa <`http://gtrc.voies.uottawa.ca/manuel/manuel-morpho-fr/index.htm`> [accessed 20 January 2011]

Martineau, France, Paul Hirschbühler, Anthony Kroch, and Yves Charles Morin. 2010. *Corpus MCVF annoté syntaxiquement*, (Ottawa: University of Ottawa) <`http://www.arts.uottawa.ca/voies/corpus_pg_en.html`> [accessed 20 January 2011]

Mathieu, Eric. 2006. 'Stylistic Fronting in Old French', *Probus*, 18: 219–66

Mazzola, Michael L. 1992. 'Stress Clash and Segment Deletion', in Laeufer and Morgan (1992), pp. 81–96

McCarthy, John. 1991. 'Synchronic Rule Inversion', in L. A. Sutton and R. Shields (eds) *The Grammar of Event Structure: General Session and Parasession, 17th Annual Meeting of the Berkeley Linguistics Society* (Berkeley: Berkeley Linguistics Society), pp. 192–207

——. 1993. 'A Case of Surface Constraint Violation', *Canadian Journal of Linguistics*, 38: 169–95

McGregor, Rob Roy (ed.). 1975. *The Lyric Poems of Jehan Froissart: A Critical Edition*, North Carolina Studies in the Romance Languages and Literatures 143 (Chapel Hill: U.N.C. Dept. of Romance Languages)

McMillan, Duncan (ed.). 1978. *Le Charroi de Nîmes: Chanson de geste du XIIe siècle, éditée d'après la rédaction AB*, 2nd rev. and corr. edn, Bibliothèque française et romane: Série B, Etudes critiques de textes 12 (Paris: Klincksieck)

Melander, Johan. 1928. *Etude sur l'ancienne abréviation des pronoms personnels régimes dans les langues romanes*, Arbeten utgivna med understöd av Vilhelm Ekmans Universitetsfond 34 (Uppsala: Almqvist & Wiksells)

Ménard, Philippe (ed.). 1987–. *Le Roman de Tristan en prose*, Textes littéraires français 353– (Geneva: Droz)

Menéndez-Pidál, Ramón. 1959. *La Chanson de Roland y el neotradicionalismo* (Madrid: Espasa-Calpe)

Mertens, Piet. 1993. 'Accentuation, intonation et morphosyntaxe', *Travaux de linguistique*, 26: 21–69

Meyer, Paul. 1892. 'L'Image du monde: Rédaction du MS Harley 4333', *Romania*, 21: 481–505

Möhren, Frankwalt (ed.). 2007. *Dictionnaire étymologique de l'ancien français: Complément bibliographique 2007* (Berlin: de Gruyter). Online version available at <`http://www.deaf-page.de/bibl_intro.htm`> [accessed 20 January 2011]

Moignet, Gérard. 1965. *Le Pronom personnel français: Essai de psycho-systématique historique*, Bibliothèque française et romane: Série A, Manuels et études linguistiques 9 (Paris: Klincksieck)

—— (ed. and trans.). 1969. *La Chanson de Roland: Texte établi d'après le manuscrit d'Oxford* (Paris: Bordas)

Mora-Lebrun, Francine (ed. and trans.). 1995. *Le Roman de Thèbes*, Lettres gothiques (Paris: Librairie générale française)

Morin, Yves-Charles. 1991. 'Old French Stress Patterns and Closed Syllable Adjustment', in Douglas A. Kibbee and Dieter Wanner (eds) *New Analyses in Romance Linguistics: Selected Papers from the XVIII Linguistic Symposium on Romance Languages, Urbana-Champaign, April 7–9, 1988*, Current Issues in Linguistic Theory 69 (Amsterdam: Benjamins), pp. 49–76

Morin, Yves-Charles, and Johnathan D. Kaye. 1982. 'The Syntactic Bases for French Liaison', *Journal of Linguistics*, 18: 291–330

Mussafia, Alfredo. 1886. 'Enclisi o proclisi del pronome personale atono quale oggeto', *Romania*, 27: 145–46

Nespor, Marina, and Irene Vogel. 1979. 'Clash Avoidance in Italian', *Linguistic Inquiry*, 10: 467–82

——. 1982. 'Prosodic Domains of External Sandhi Rules', in Harry van der Hulst and Norval Smith (eds) *The Structure of Phonological Representations*, Linguistic models (Dordrecht: Foris), pp. 225–65

——. 1986. *Prosodic Phonology* (Dordrecht: Foris)

——. 1989. 'On Clashes and Lapses', *Phonology*, 6: 69–116

Nicols, Stephen G., Jr. 1963. 'Style and Structure in *Gormont et Isembart*', *Romania*, 54: 500–535

Noomen, Willem, and Nico van den Boogaard (eds). 1983–98. *Nouveau Recueil complet des fabliaux (NRCF)*, 10 vols (Assen: Van Gorcum)

Noyer, Rolf. 2002. 'Generative Metrics and Old French Octosyllabic Verse', *Language Variation and Change*, 14: 119–71

Nykrog, Per. 1973. *Les Fabliaux*, new edn (Geneva: Droz)

Offord, Malcolm H. 1971. 'The Use of Personal Pronoun Subjects in Post-Position in Fourteenth Century French', *Romania*, 92: 37–64, 200–45

Palmer, R. Barton (ed.). 1984. *The Judgment of the King of Bohemia = Le Jugement dou roy de Behaingne*, by Guillaume de Machaut, Garland Library of Medieval Literature Series A 9 (New York: Garland)

Palsgrave, John. 1530. *L'Esclarcissement de la langue francoyse* ([London]: [Richard Pynson,] Johan Haukyns). Facsimile reprint with French translation and notes by Susan Baddeley. 2003. (Paris: Champion)

Paris, Gaston. 1902. Review of Kerstin Hård: 'Sur l'âge de l'auteur du fragment de Brussels, Gormund et Isembard', *Romania*, 31: 445–48

Paris, Gaston, and Léopold Pannier (eds). 1887. *La Vie de saint Alexis: Poème du XIe siècle et renouvellements des XIIe, XIIIe et XIVe siècles* (Paris: F. Vieweg)

Paris, Gaston, and Ulysse Robert (eds). 1876–93. *Miracles de Nostre Dame par personnages*, 8 vols, Société des anciens textes français (Paris: Firmin Didot)

Pasdeloup, Valérie. 1992. 'A Prosodic Model for French Text-to-Speech Synthesis: A Psycholinguistic Approach', in Gérard Bailly and Christian Benoit (eds) *Talking Machines: Theories, Models, and Designs* (Amsterdam; London: North-Holland), pp. 335–48

Pauphilet, Albert (ed.). 1923. *La Queste del saint Graal: Roman du XIIIe siècle*, Classiques français du Moyen Age 33 (Paris: Champion)

Pensom, Roger. 1982. *Literary Technique in the Chanson de Roland*, Histoire des idées et critique littéraire 203 (Geneva: Droz)

——. 1985. 'On the Prosody of the Decasyllabic Lyrics of the 'Roi de Navarre'', *French Studies*, 39: 257–75

——. 2000. *Accent and Metre in French: A Theory of the Relation between Linguistic Accent and Metrical Accent in French, 1100–1900* (Bern: Peter Lang)

——. 2004. *Le Sens de la métrique chez François Villon: Le Testament* (Bern: Peter Lang)

——. 2009. 'Accent et syllabe dans les vers français: Une Synthèse possible?', *Journal of French Language Studies*, 19: 335–63

Peperkamp, Sharon. 2004. 'Lexical Exceptions in Stress Systems: Arguments from Early Language Acquisition and Adult Speech Perception', *Language*, 80: 98–126

Peperkamp, Sharon, and Emmanuel Dupoux. 2002. 'A Typological Study of Stress "Deafness"', in Gussenhoven and Warner (2002), pp. 203–40

Peperkamp, Sharon, Emmanuel Dupoux, and Núria Sebastián-Gallés. 1999. 'Perception of Stress by French, Spanish, and Bilingual Subjects', in *Proceedings, Sixth European Conference on Speech Communication and Technology (EUROSPEECH '99)*, Budapest, Hungary, September 5–9, 1999, pp. 2683–2686

Perry, Anne Joubert Amari (ed.). 1981. *La Passion des jongleurs: Texte établi d'après la Bible des sept estaz du monde de Geufroi de Paris*, Textes, dossiers, documents 4 (Paris: Beauchesne)

Picot, Emile (ed.). 1902–12. *Recueil général des sotties*, 3 vols, Société des anciens textes français (Paris: Firmin Didot)

Poletto, Cecilia. 2005. '*Sì* and *e* as CP Expletives in Old Italian', in Batllori and Roca (2005), pp. 206–35

Pope, Mildred Katharine. 1918. 'The Dialect of "Gormont and Isembart"', *The Modern Language Review*, 13: 335–38

——. 1952. *From Latin to Modern French with Especial Consideration of Anglo-Norman*, 2nd edn (Manchester: Manchester University Press)

Post, Brechtje. 2000. *Tonal and Phrasal Structures in French Intonation*, LOT international series (The Hague: Thesus)

Prévost, Sophie. 2001. 'Etiquetage morpho-syntaxique de la BFM: CATTEX', unpublished manuscript, ENS-LSH

Price, Glanville. 1998. *The French Language: Present and Past*, revised and updated edn (London: Grant & Cutler)

Priestley, L. 1955. 'Reprise Constructions in French', *Archivum Linguisticum*, 7: 1–28

Pulgrum, Ernst. 1965. 'Prosodic Systems: French', *Lingua*, 13: 125–44

Rainsford, Thomas M. forthcoming. 'Dividing Lines: The Changing Syntax and Prosody of the Mid-Line Break in Medieval French Octosyllabic Verse', *Transactions of the Philological Society*

Ramsden, Herbert. 1963. *Weak-Pronoun Position in the Early Romance Languages* (Manchester: Manchester University Press)

Ramus, Franck, Marina Nespor, and Jacques Mehler. 1999. 'Correlates of Linguistic Rhythm in the Speech Signal', *Cognition*, 72: 1–28

Raynaud de Lage, Guy (ed.). 1966. *Le Roman de Thèbes*, 2 vols, Classiques français du Moyen Age 94, 96 (Paris: Champion)

—— (ed.). 1976. *Eracle*, by Gautier d'Arras, Classiques français du Moyen Age 102 (Paris: Champion)

Reenan, Pieter van, and Lene Schøsler. 1992. 'Ancien et moyen français: *Si* 'thématique', analyse exhaustive d'une série de textes', *Vox Romanica*, 51: 101–27

——. 1993. '*Si* 'thématique': Etude de *si* en ancien et moyen français, discours direct', in Gerold Hilty (ed.) *Actes du XXe Congrès International de Linguistique et Philologie Romanes: Université de Zurich (6–11 avril 1992)*, 5 vols (Tübingen: Francke), vol. 1, pp. 617–28

——. 1995. 'The Thematic Structure of the Main Clause in Old French: *Or* versus *si*', in Henning Andersen (ed.) *Historical linguistics 1993: Selected Papers from the 11th International Conference on Historical Linguistics, Los Angeles, 16–20 August 1993*, Current Issues in Linguistic Theory 124 (Amsterdam: Benjamins), pp. 401–19

——. 2000. 'The Pragmatic Function of the Old French Particles AINZ, APRES, DONC, LORS, OR, PUIS, SI', in Susan C. Herring, Pieter van Reenan, and Lene Schøsler (eds) *Textual Parameters in Older Languages*, Current Issues in Linguistic Theory 195 (Amsterdam; Philadelphia: Benjamins), pp. 59–106

Reno, Christine, and Liliane Dulac (eds). 2001. *Le Livre de l'advision Cristine*, by Christine de Pisan, Etudes christiniennes 4 (Paris: Champion)

Richter, Elise. 1903. *Zur Entwicklung der romanischen Wortstellung aus dem lateinischen* (Halle a. S.: Niemeyer)

—— (ed.). 1914. *Aeneas Sylvius Eurialus und Lukrezia: übersetzt von Octovien de Saint-Gelais, nebst bruchstücken der Anthitus-übersetzung* (Halle: Niemeyer)

Riegel, Martin, Jean-Christophe Pellat, and René Rioul. 1994. *Grammaire méthodique du français* (Paris: Presses Universitaires de France)

Rizzi, Luigi. 1997. 'The Fine Structure of the Left Periphery', in Liliane Haegeman (ed.) *Elements of Grammar: Handbook in Generative Syntax* (Dordrecht: Kluwer), pp. 281–337

Rizzi, Luigi (ed.). 2004. *The Structure of CP and IP: The Cartography of Syntactic Structures, vol. 2* (Oxford: Oxford University Press)

Roach, William (ed.). 1959. *Le Roman de Perceval ou le conte du Graal: Publié d'après le ms fr. 12576 de la Bibliothèque Nationale*, by Chrétien de Troyes, 2nd rev. and expanded edn, Textes littéraires français 71 (Geneva: Droz; Paris: Minard)

Robert, Ulysse (ed.). 1897. *Li Abrejance de l'ordre de chevalerie*, by Jean Priorat de Besançon, Société des anciens textes français (Paris: Firmin Didot)

Roberts, Ian G. 1993. *Verbs and Diachronic Syntax: A Comparative History of English and French*, Studies in Natural Language and Linguistic Theory 28 (Dordrecht: Kluwer)

——. 2004. 'The C-system in Brythonic Celtic Languages: V2, and the EPP', in Rizzi (2004), pp. 297–328

Robson, Charles Alan. 1961. 'Les Origines de la langue littéraire en France: Rime approximative et assonance', *Boletim de Filolgía*, 19: 11–27

Rochet, Bernard L. 1976. *The Formation and Evolution of the French Nasal Vowels*, Beihefte zur Zeitschrift für romanische Philologie 153 (Tübingen: Niemeyer)

Roques, Mario (ed.). 1931. *Le Roman du comte d'Anjou*, by Jehan Maillart, Classiques français du Moyen Age 67 (Paris: Champion)

—— (ed.). 1957. *L'Estoire de Griseldis: En rimes et par personnages (1395)*, Textes littéraires français 74 (Geneva: Droz)

—— (ed.). 1958. *Romans de Chrétien de Troyes, édités d'après la copie de Guiot (Bibl. nat. fr. 794), vol. 3: Le Chevalier de la charrete*, Classiques français du Moyen Age 86 (Paris: Champion)

Rosenberg, Samuel N., and Samuel Danon (eds and trans.). 1985. *The Lyrics and Melodies of Gace Brulé*, Garland Library of Medieval Literature Series A 39 (New York: Garland)

Rossi, Mario. 1980. 'Le Français, langue sans accent?', in Fónagy and Léon (1980), pp. 13–51

——. 1985. 'L'Intonation et l'organisation de l'énoncé', *Phonetica*, 42: 135–53

Roussel, Henri (ed.). 1961. *Renart le nouvel*, by Jacquemars Giélée, Société des anciens textes français (Paris: A. & J. Picard)

Ruelle, Pierre (ed.). 1960. *Huon de Bordeaux*, Travaux de la Faculté de philosophie et lettres 20 (Brussels: Presses universitaires)

Runnalls, Graham A (ed.). 1995. *Le Mystère de Judith et Holofernes: Une Edition critique de l'une des parties du "Mistere du Vieil Testament"*, attributed to Jean Molinet, Textes littéraires français 461 (Geneva: Droz)

Russell, Delbert W. (ed.). 1995. *La Vie seint Richard evesque de Cycestre*, by Pierre D'Abernon of Fetcham, Anglo-Norman Texts 51 (London: Anglo-Norman Text Society)

Rychner, Jean, and Albert Henry (eds). 1974. *Le Testament Villon*, Textes littéraires français 208 (Geneva: Droz)

—— (eds). 1977. *Le Lais Villon et les poèmes variés*, by François Villon, 2 vols, Textes littéraires français 239, 240 (Geneva: Droz)

Rydberg, Gustaf. 1896–1907. *Geschichte des Französischen ə*, 6 vols (Uppsala: Almqvist & Wiksells)

Santorini, Beatrice. 2010. 'Annotation Manual for the Penn Historical Corpora and the PCEEC', technical manual, University of Pennsylvania <`http://www.ling.upenn.edu/~beatrice/annotation/`> [accessed 20 January 2011]

Schmid, Helmut. 1994. *TreeTagger* (Stuttgart: University of Stuttgart) [software] <`http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/`> [accessed 20 January 2011]

Segre, Cesare (ed.). 1989. *La Chanson de Roland*, editorial material translated by Madeleine Tyssens, 2 vols, 2nd edn, Textes littéraires français 368 (Geneva: Droz)

Selkirk, Elisabeth. 1981. 'On Prosodic Structure and its Relation to Syntactic Structure', in Thorstein Fretheim (ed.) *Nordic Prosody II* (Trondheim: TAPIR), pp. 111–40

——. 1984. *Phonology and Syntax: The Relation Between Sound and Structure*, Current studies in linguistics series (Cambridge, Mass: MIT Press)

——. 1986. 'On Derived Domains in Sentence Phonology', *Phonology Yearbook*, 3: 371–405

——. 1996. 'The Prosodic Structure of Function Words', in James L. Morgan and Katherine Demuth (eds) *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (Mahwah, N.J.: Erlbaum), pp. 187–214

Selkirk, Elisabeth, and Tong Shen. 1990. 'Prosodic Domains in Shanghai Chinese', in Sharon Inkelas and Draga Zec (eds) *The Phonology-Syntax Connection* (Chicago: University of Chicago Press), pp. 313–38

Shattuck-Hufnagel, Sharon. 1991. 'Acoustic Correlates of Stress Shift', in *Actes du XIIe congrès international des sciences phonétiques: 19–24 août 1991, Aix-en-Provence, France*, 5 vols (Aix-en-Provence: Université de Provence), vol. 4, pp. 266–69

Short, Ian (ed.). 2005. 'Part 1: The Oxford Version', in Joseph J. Duggan (ed.) *La Chanson de Roland = The Song of Roland: The French Corpus*, 3 vols (Turnhout: Brepols), vol. 1, pp. I/1–I/338

Short, Ian, and Brian Merrilees (eds). 1979. *The Anglo-Norman Voyage of St. Brendan*, by Benedeit (Manchester: Manchester University Press)

Skårup, Povl. 1975. *Les Premières Zones de la proposition en ancien français: Essai de syntaxe de position*, Revue Romane numéro spécial 6 (Copenhagen: Akademisk forlag)

Solente, Suzanne (ed.). 1959–66. *Le Livre de la mutacion de fortune*, by Christine de Pisan, 4 vols, Société des anciens textes français (Paris: A. & J. Picard)

Speyer, Augustin. 2010. *Topicalization and Stress Clash Avoidance in the History of English*, Topics in English Linguistics 69 (Berlin; New York: De Gruyter)

Stebbins, Charles Ernest. 1971. 'The Oxford Version of the Vie de saint Alexis: An Old French Poem of the Thirteenth Century', *Romania*, 92: 1–36

Stein, Achim. 2007. 'Resources and Tools for Old French Text Corpora', in Yuji Kawaguchi and Yoichiro Tsuruga (eds) *Corpus-Based Perspectives in Linguistics*, Usage-Based Linguistic Informatics 6 (Amsterdam: Benjamins), pp. 217–29

Stein, Achim, and Martin-Dietrich Gleßgen. 2005. 'Resources and Tools for Analyzing Old French Texts', in Claus D. Pusch and Wolfgang Raible (eds) *Romanistische Korpuslinguistik II: Korpora und diachrone Sprachwissenschaft = Romance Corpus Linguistics II: Corpora and Diachronic Linguistics* (Tübingen: Narr), pp. 135–45

Stein, Achim, Pierre Kunstmann, and Martin-Dietrich Gleßgen (eds). 2006. *Nouveau Corpus d'Amsterdam: Corpus informatique de textes littéraires d'ancien français (ca 1150–1350), établi par Anthonij Dees (Amsterdam 1987)* (Stuttgart: Institut für Linguistik / Romanistik)

Storey, Christopher (ed.). 1968. *La Vie de saint Alexis: Texte du manuscrit de Hildesheim (L)*, Textes littéraires français 148 (Paris: Minard)

Suchier, Hermann. 1878. 'Die Mundart des Leodegarliedes', *Zeitschrift für romanische Philologie*, 2: 255–302

Tarlinskaja, Marina, and L. M. Teterina. 1974. 'Verse—Prose—Meter', *Linguistics*, 129: 63–86

Taylor, Ann, Arja Nurmi, Anthony Warner, Susan Pintzuk, and Terttu Nevalainen. 2006. *Parsed Corpus of Early English Correspondence, Parsed Version* (York: University of York; Helsinki: University of Helsinki). Compiled by the CEEC Project Team, distributed through the Oxford Text Archive.

Taylor, Ann, Anthony Warner, Susan Pintzuk, and Frank Beths. 2003. *York-Toronto-Helsinki Parsed Corpus of Old English Prose*, (York: University of York) <`http://www-users.york.ac.uk/~lang22/YCOE/YcoeHome.htm`> [accessed 20 January 2011]

Thiersch, Craig Lee. 1978. 'Topics in German Syntax', unpublished doctoral dissertation, M.I.T.

Thurneysen, Rudolf. 1892. 'Die Stellung des Verbums in Altfranzösischen', *Zeitschrift für romanische Philologie*, 16: 289–307

Tobler, Adolf. 1875. 'Besprechung von J. Le Coultre, *De l'ordre des mots dans Chrétien de Troyes*', *Vermischte Beiträge zue französischen Grammatik*, 5: 395–414

Tobler, Adolf, and Erhard Lommatzsch (eds). 1925–. *Altfranzösisches Wörterbuch* (Berlin [1925–36]: Weidmannsche Buchhandlung; Wiesbaden [1951–76]: F. Steiner; Stuttgart [1989–]: F. Steiner)

Tranel, Bernard. 1987. *The Sounds of French: An Introduction* (Cambridge: Cambridge University Press)

Tremblay, Annie. 2009. 'Phonetic Variability and the Variable Perception of L2 Word Stress by French Canadian Listeners', *International Journal of Bilingualism*, 13: 35–62

Urwin, Kenneth (ed.). 1958. *Le Lyon coronné (1467): Texte bourguignon inédit*, Textes littéraires français 81 (Geneva: Droz)

Väänänen, Veikko. 1981. *Introduction au latin vulgaire*, 3rd rev. and expanded edn, Bibliothèque française et romane: Série A, Manuels et études linguistiques 6 (Paris: Klincksieck)

Vance, Barbara. 1988. 'Null Subjects and Syntactic Change in Medieval French', unpublished doctoral dissertation, Cornell University

——. 1995a. 'On the Clitic Nature of Subject Pronouns in Medieval French', in A. Dainora, R. Hemphill, B. Luka, B. Need, and S. Pargman (eds) *Papers from the 31st Meeting of the Chicago Linguistic Society*, 2 vols (Chicago, Ill.: Chicago Linguistic Society), vol. 2, pp. 300–315

——. 1995b. 'On the Decline of Verb Movement to Comp in Old and Middle French', in Battye and Roberts (1995), pp. 173–199

——. 1997. *Syntactic Change in Medieval French: Verb-Second and Null Subjects*, Studies in Natural Language and Linguistic Theory 41 (Dordrecht: Kluwer)

Varty, Kenneth (ed.). 1965. *Christine de Pisan's Ballades, Rondeaux and Virelais: An Anthology* (Leicester: Leicester University Press)

Vennemann, Theo. 1974. 'Topics, Subjects and Word Order', in J. M. Anderson and Charles Jones (eds) *Historical Linguistics: Proceedings of the First International Conference on Historical Linguistics, Edinburgh, 2nd–7th Sept. 1973*, 2 vols (Amsterdam: North Holland), vol. 1, pp. 339–76

Verluyten, Sylvain Paul. 1982. 'Recherches sur la prosodie et la métrique de français', unpublished doctoral thesis, University of Antwerp

——. 1984. 'Phonetic Reality of Linguistic Structures: The Case of (Secondary) Stress in French', in M. P. R. van den Broecke and Antonie Cohen (eds) *Proceedings of the Tenth International Congress of Phonetic Sciences*, Netherlands Phonetic Archives 2B (Dordrecht: Foris), pp. 522–26

Vising, Johan. 1923. *Anglo-Norman Language and Literature* (Oxford: Oxford University Press)

Vitz, Evelyn Birge. 1999. *Orality and Performance in Early French Romance* (Cambridge: D. S. Brewer)

Vogel, Irene, Timothy Bunnell, and Steven Hoskins. 1995. 'The Phonology and Phonetics of the Rhythm Rule', in Connell and Arvaniti (1995), pp. 111–27

Wallensköld, Axel (ed.). 1921. *Les Chansons de Conon de Bethune*, Classiques français du Moyen Age 24 (Paris: Champion)

—— (ed.). 1925. *Les Chansons de Thibaut de Champagne, roi de Navarre*, Société des anciens textes français (Paris: Champion)

Walter, Henriette. 1982. *Enquête phonologique et variétés régionales du français* (Paris: Presses Universitaires de France)

Wooldridge, Jeffrey M. 2006. *Introductory Econometrics: A Modern Approach*, 3rd edn (Mason, OH: Thomson Higher Education)

Zink, Gaston. 1992. *Morphologie du français médiéval*, 2nd rev. edn (Paris: Presses Universitaires de France)

Zink, Michel. 1995. *Medieval French Literature: An Introduction*, translated by Jeff Rider, Medieval & Renaissance Texts & Studies 110 (Binghamton, NY: Medieval & Renaissance Texts & Studies)

Zumthor, Paul. 1972. *Essai de poétique médiévale* (Paris: Seuil)