Anchorites and abbreviations: a corpus study of abbreviations of Germanic and Romance lexicon in the *Ancrene Wisse*

**Abstract:** Manuscript abbreviations are a well-known feature of manuscript culture, which have mainly been studied qualitatively by palaeographers. The present study uses a quantitative corpus-based approach to examine how abbreviations are distributed in the etymologically Romance and Germanic lexicons during the early Middle English period (1150–1350), which saw many developments in the writing systems of English. It applies linear regression with effects coding on a dataset consisting of all the versions of *Ancrene Wisse* included in the *Linguistic Atlas of Early Middle English (LAEME)*. The results reveal a statistically significant distribution in which some abbreviations are used exclusively for Germanic words, some for Romance and some for both, proving that ‘Romance’ and ‘Germanic’ work as diagnostic categories. Further corpus searches reveal a group of abbreviations that are used almost exclusively in the West Midlands.

**Keywords:** corpus linguistics, manuscript abbreviations, *Ancrene Wisse*, lexicology, linear regression, historical dialectology, early Middle English, *LAEME*

1. Introduction

The application of quantitative methods to digitized medieval data has opened up new prospects for data mining. One largely unexplored frontier is the various medieval abbreviation and suspension characters that were used to save parchment, paper or time. The current paper is a pilot study which tests how well ‘Romance’ and ‘Germanic’ work as diagnostic categories to explain how different types of abbreviations are distributed in seven samples of the work known as *Ancrene Wisse* or *Ancrene Riwle*.\(^1\) It is part of a larger project called *A Corpus Approach to Manuscripts (CAM)*.\(^2\) The project aims to investigate the distribution of abbreviations in the *Linguistic Atlas of Early Middle English (LAEME)*, using quantitative corpus- and statistics-based methodologies.

Abbreviations are a well-known feature of handwritten book culture, and they can be used to aid dating and identification of scribal hands. For example, the prominent German palaeographer Ludwig Traube wrote in 1902: “when I am inquiring into the date of a manuscript,

---

\(^1\) Somewhat confusingly, the same work is known by two titles: *Ancrene Riwle* and *Ancrene Wisse*. I follow the practice of Millett (2005) and Millett and Dance (2006) of referring to the entire work across all manuscript witnesses as *Ancrene Wisse*, because the title is medieval. It is found in the manuscript Cambridge, Corpus Christi College 402. The title *Ancrene Riwle*, which has gained some currency, was coined by James Morton (1853). It is not a bad proposal, as it is influenced by Latin titles such as *Regula Sancti Benedicti* ‘the Rule of Saint Benedict’ and since words *ancrene* and *riwle* are attested several times early in the work. Nevertheless, since this is one of the rare occasions on which a medieval title survives and more recent scholarship prefers *Ancrene Wisse*, I will also use it.

\(^2\) The project is funded by the Swiss National Foundation (SNF) for the period 2017–2020. I am grateful for their support. I am also grateful to two Edinburgh colleagues, Daisy Smith, who helped me with statistics, and Raffaela Baechler, who helped me with French, as well as to the two peer reviewers, Keith Williamson and Merja Stenroos. Without their suggestions this would be a much worse article. Needless to say, all remaining mistakes are my own.
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinnen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. Current Explorations in Middle English. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

I immediately turn to the abbreviations” (translated by Clemens and Graham 2007: 89).³ It is rather logical to hypothesize that abbreviations could be used as data for quantitative analysis, which could be used in historical linguistics and as a help for identifying scribes. However, as digital data, abbreviations have so far been barely mined: we do not know which variables should be examined and there are entire periods which have not been studied.⁴ The current study aims to address the problem of examining the distribution of abbreviations in the etymological Romance and Germanic lexicons during the formative years of the early Middle English period.

The medieval abbreviation and suspension system developed “to shorten the labour of writing Latin” (Hector 1958: 37) over many centuries of writing texts to parchment or papyrus, or making quick transcriptions from speech. When vernacular languages began to be written down, the abbreviation system was adapted to writing them with some modifications, such as the Old English and Old Norse abbreviations involving the runic letter thorn (cf. West 2006). The system was also adapted to writing Romance vernaculars, and application was, in some respects, more straightforward (Hector 1958: 36).⁵ For example, four Latin p-abbreviations, p ‘per’, p⁹ ‘prae’, p‘ pro’, and p₉ ‘post’, which had developed in the early minuscule period (Hasenohr 2002: 80), could be directly used to abbreviate French ‘par’ (on occasion ‘per’), ‘pré’, ‘pro’, and ‘quis’ (Hector 1958: 36, see also Hasenohr 2002: 86). Some Latin contractions were also easily applied to equivalent French words, n’re ‘nôstre’ and b’n ‘bien’ (cf. Hector 1958: 36, Johnson and Jenkinson 1915: xxvii), and superscripts are attested too, including g’t ‘grant’ and q’t ‘que’ (Hasenohr 2002: 86). Consequently, it is of interest to study how abbreviations were distributed in the Romance and Germanic vocabulary of Middle English, especially in the period after the Norman Conquest, which saw many changes both to the writing system and the vocabulary of Middle English.

The present study focuses on the early Middle English period (1150–1350), which is an especially interesting time for the development of the English writing system. The Norman Conquest of 1066 introduced a new ruling class and relegated English to a tertiary role as a written language after Latin and Anglo-Norman French. Scribal practices in the period show “a great deal of experimentation,” as the scribes’ “attempts to represent their native language include the entire historical repertoire (Anglo-French and Anglo-Latin as well as Old English traditions) adapted with varying degrees of individual inventiveness” (Laing and Lass 2006: 17.2.2.). Consequently, it is of interest to see, whether there is a difference in their abbreviation practices for Germanic and Romance vocabulary.

In the study, I will examine the following research questions:

1. What is the proportion of abbreviated words of the full word count?

2. Which words get abbreviated more frequently: Romance or Germanic?

³ The original reads: ‘Ego cum aetatem codicis sciscitor, statim me ad compendia verto’. (Traube 1902 as quoted by Clemens & Graham 2007: 89)
⁴ There have been a handful studies. Shute (2017) studied abbreviations as a feature of line justification in early printed books, noting that they are statistically more likely to occur close to the right margin. Smith (2016) studied abbreviations in Older Scots, and discovered that the ¹ abbreviation (for ‘er’, ‘e’ or final ‘-e’) is more likely to occur after certain letter shapes. Other discoveries were made by Rogos (2012), who focused on late Middle English literary manuscripts, noting that word-final characters alternate with graphic sequences rather than substitute them. Honkapohja (2018) presented a pilot study of abbreviations in Latin and Middle English in the Voigts-Sloane Group of manuscripts, focusing on content and function words. Thus, the existing studies have uncovered that abbreviations can be conditioned by the context of the page or preceding graph, and suggested possibilities for using abbreviations in authorship attribution and the study of multilingual practices.
⁵ Abbreviations were not as frequent in Romance as in Latin though. Hector notes that “Anglo-Norman documents are very much less heavily abbreviated than the Latin ones contemporary with them” (Hector 1958: 36). Hasenohr (2002: 82–83) also notes in her calculations much fewer abbreviations in French compared to contemporary Latin.
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. Current Explorations in Middle English. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

3. Do certain abbreviations occur more frequently with etymologically Romance or Germanic vocabulary?

The methodology consists of analysing a linguistic corpus through functions offered by the statistical programming language R, using linear regression and effects coding, using Romance and Germanic as well as abbreviation types as independent variables and token count as the dependent variable. I also made use of the mapping feature of LAEME to investigate the geographical spread of certain abbreviations and abbreviation practices.

2. The Ancrene Wisse corpus

Another reason why the early ME period is a good one to focus on is that a suitable resource is now available for its study. The data in the present study are collected from the Linguistic Atlas of Early Middle English (LAEME) corpus, compiled by Margaret Laing and first published in 2008. It is based on a philosophy according to which written language should be examined in its own right (McIntosh 1956), not only as a reflection of spoken language (Laing 2000: 98–100). The genres in LAEME include official records, prose, poetry and lyrics. The length of the texts in the LAEME corpus varies between 18 and 30,500 words. Altogether the corpus is ca. 650,000 words in size. The following reasons make the corpus particularly well-suited for the present study.

A) Each text in LAEME is based on a diplomatic transcription from manuscript facsimiles, not editions. Its collection principles avoid editorial interpretation such as silently expanding abbreviations and correcting perceived mistakes (LAEME 3.3.1).

B) The corpus is specifically designed for dialectological analysis. It includes altogether 167 scribal texts, or using the LAEME term, “specimens”, of early Middle English, representing language written down between 1150 and 1350. Each specimen represents the language of one scribe, meaning that scribal stints are clearly differentiated (LAEME 3.2.).

C) The specimens were collected and localised based on the approach to historical dialectology developed earlier for the LALME, including anchor texts localised by external information and others localised in relation to the anchor texts, using the ‘fit’ technique (cf. Laing and Lass 2006: 17.2.1.1)

D) The corpus includes 32 texts in parallel versions, that is, copies of the same text by different scribes. Shorter texts are included in their entirety and longer texts as corresponding samples of the same passage to make comparison possible (LAEME 3.1.).

The data included in this pilot study consist of the seven scribal ‘specimens’ of Ancrene Wisse which are included in LAEME. As two manuscripts are represented by samples taken from two scribes or annotators, altogether these specimens come from five manuscripts. I decided to focus on the AW, because it is one of the works available in parallel versions in LAEME and also because the physical manuscripts are fairly small books, designed for private devotion. My previous work suggests that small size corresponds with greater use of abbreviations (cf. Honkapohja 2018: 249–251).

The AW is a collection of advice written for anchoresses, female religious recluses, by an anonymous author. It has been described as “the longest, the most complex, and the most influential of an unexpectedly sophisticated group of religious prose works, produced in the
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thangs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

West Midlands at a time when relatively little was being written in Middle English” (Millett and Dance 2006: ix). The work survives in several manuscripts, including early manuscripts from the 1230s or 1240s, such as Cotton Nero and Cotton Cleopatra, and also later ones such as a fourteenth-century version in the Vernon Manuscript (Oxford, Bodleian Library, MS. Eng. Poet. a. 1), a translation into Anglo-Norman French and early printed editions, such as *Tretysye of Love*, printed by Wynkyn de Worde in 1493/4 (STC (2nd ed) / 24234, ed. Fisher 1951).

The *LAEME* data are drawn from the five surviving manuscripts of the work which can be dated to the period of the corpus (1150–1350). These include three textually close versions found, respectively, in London, British Library, Cotton Cleopatra C vi, Cambridge, Corpus Christi College 402 and London, British Library, Cotton Nero A, which are transcribed in samples using “corresponding portions to make close comparison possible” (*LAEME* 3.1). In addition, *LAEME* also includes a second specimen from the Cotton Cleopatra C vi, comprising of corrections made to the work of the main scribe by a second hand, which may very well belong to the author of the treatise (cf. Dobson 1972, 1976).

The different copies of AW can differ from each other considerably, as copyists appear to have adapted the text to the personal, social and geographical needs of a changing audience (cf. Doyle 1954: 234; Millett 2005: xxxvii). *LAEME* includes one specimen from Cambridge, Gonville and Caius College 234/120, which is a “much shortened and reordered version”. This copy also systematically changes the frequently appearing addresses to mine leoue sustren ‘my beloved sisters’ to brepren – suggesting that it was adapted to a male community (Millett 2005: xvi).

The fifth manuscript included in *LAEME*, London, British Library, Cotton Titus D xviii, is considered to be problematic by the compiler of *LAEME*, because it, in her interpretation, represents a mixed language, produced by a partly *litteratim* copying scribe, who copied some forms in the exemplar and replaced others with his own dialect. For this reason, the Titus version is represented by two samples, one comprising “a layer of consistent homogeneous usage” in a North West Midland type of language (*titusar*) (*LAEME* 3.1). The second is a mixed-language specimen (*tituslang2*), which is a mixture of the dialect of the first sample and “something more southerly” (*LAEME*: Index of Sources).

Table 1 presents the file name, date, manuscript shelf-number, word-count and abbreviation count of the corpus used in the present study.

Table 1. The *Ancrene Wisse* corpus.

<table>
<thead>
<tr>
<th>Specimen</th>
<th><em>LAEME</em> ID</th>
<th>Date</th>
<th>Library shelf mark</th>
<th>Words</th>
<th>Abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>cleoara</td>
<td>#273</td>
<td>1230s</td>
<td>London, British Library, Cotton Cleopatra C vi</td>
<td>14,033</td>
<td>1,628</td>
</tr>
<tr>
<td>cleoarb</td>
<td>#275</td>
<td>1230</td>
<td>London, British Library, Cotton Cleopatra C vi, annotations</td>
<td>1,047</td>
<td>106</td>
</tr>
<tr>
<td>titusar</td>
<td>#118</td>
<td>1240–1250</td>
<td>London, British Library, Cotton Titus D xviii, sample fitted to Cheshire</td>
<td>14,224</td>
<td>1,431</td>
</tr>
<tr>
<td>tituslang2</td>
<td>#119</td>
<td>1240–1250</td>
<td>London, British Library, Cotton Titus D xviii, sample containing mixed language</td>
<td>7,436</td>
<td>1,233</td>
</tr>
<tr>
<td>neroar</td>
<td>#245</td>
<td>1240s</td>
<td>London, British Library, Cotton Nero A xiv</td>
<td>15,285</td>
<td>1,693</td>
</tr>
</tbody>
</table>
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thensg and Oliver M. Traxel (eds.). 2019. Current Explorations in Middle English. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

<table>
<thead>
<tr>
<th></th>
<th>#</th>
<th>1250–1275</th>
<th>Cambridge, Gonville and Caius College 234/120</th>
<th>8,845</th>
<th>345</th>
</tr>
</thead>
<tbody>
<tr>
<td>caiusar</td>
<td>#276</td>
<td>1250–1275</td>
<td>Cambridge, Gonville and Caius College 234/120</td>
<td>8,845</td>
<td>345</td>
</tr>
<tr>
<td>corpar</td>
<td>#272</td>
<td>1270s or early 1280s</td>
<td>Cambridge, Corpus Christi College 402</td>
<td>15,183</td>
<td>1,095</td>
</tr>
</tbody>
</table>

3. Methodology

This study combines methods from corpus linguistics, statistical analysis, and historical dialectology. First, I performed corpus searches using the AntConc software and LAEME .dic and .txt files to find all abbreviations and compiled word lists of abbreviated words for each specimen based on the lexels of LAEME. Second, based on the word lists, I compiled a dataset of these abbreviations, which includes:

a) specimen name
b) token frequency of abbreviation types
c) classification into Romance and Germanic words
d) abbreviation type

Third, the dataset was examined by preparing descriptive statistics and using the statistical programming language R to test whether a model in which abbreviation types and the classification into etymologically Romance and Germanic words are used as independent variables and token frequencies as dependent variables would produce statistically significant results. Finally, I also used the mapping tool in LAEME to examine the regional distribution of abbreviated forms.

Compiling the dataset required decisions some of which need to be explained in more detail. Category a) was described in section 2 above. Category b) is fairly self-evident: it is the token frequencies of abbreviation types in the dataset. However, the decisions made for the independent variables c) Romance vs. Germanic and d) abbreviation type require explication.

3.1 Romance vs. Germanic

The main research question which I am addressing in this study is whether classifying all abbreviated words based on etymology into two groups, ‘Romance’ and ‘Germanic’, will uncover significant variation in use within the two classes of words. As the lexicon of any language is unavoidably heterogeneous, dividing it into two categories is not without its problems and requires making decisions, which I will discuss here.

First, Romance influence on English came over a long period of time and led to the adoption of a very large and heterogeneous group of words (see e.g. Kastovsky 1992). Etymologically, Romance words range from early loans from classical Latin, which were borrowed when Germanic tribes came into contact with the Roman empire (cat, cheese, street), to recent borrowings from Anglo-Norman French (prison, spouse). When it comes to the domain of religious vocabulary attested in the AW, it would be possible to argue that some Latin borrowings are so fully integrated that they are more Germanic than Romance. This includes a word such as priest (Old English: preost, etymon: post-classical Latin presbyter), which was borrowed around the time of the conversion of the Anglo-Saxons, is fully integrated into the
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. Current Explorations in Middle English. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

sound system and has cognates in most Germanic languages. Nevertheless, it would be difficult to come up with suitable criteria for determining the level of integration given the heterogeneity of vocabulary. Classifying Latin words that occur already in Old English as Germanic would also include words which clearly do represent Latin written culture and abbreviation practices, for example, crist ‘christ’, which could be abbreviated either using superscript c'ist or a nomen sacrum: xps. I therefore decided to consider all words defined in the OED and MED as borrowings from Romance languages, irrespective of the period of borrowing, as Romance words.

I also classified as Romance words originally borrowed into Latin from other languages, such as Greek (martyr), Hebrew (amen), or Etruscan (person, possibly from Etruscan ϕερσu ‘mask’), as all of them were borrowed into English via Latin and represent the type of Christian Latin traditions which I aimed to separate from ones developed for writing Germanic languages. Moreover, the personal name Jesus is counted as a Romance loan, since it was borrowed into English via Latin.

Germanic, on the other hand, denotes mainly words descended from Old English of proto-Germanic origin. Scandinavian borrowings are not separated from the rest of the Germanic vocabulary, including such recent items as to mistake (take = a Scandinavian borrowing, mis- a common Germanic prefix). However, there were very few Scandinavian words in the data.

3.2 Abbreviation types

Another feature of the dataset compilation that required separate decisions was classifying abbreviations. I mainly follow the decisions made in the compilation of LAEME (see introduction chapter 3 for a discussion). As the number of encoded features was already high, on a few occasions I combined separately encoded features into a single category, most notably in the case of superscript abbreviations, which is an open-ended productive category. The categories used in the present study are displayed in Table 2.

---

6 Although according to OED “the phonology is difficult to explain”.
7 According to OED, the earliest English attestation of the word person is in AW. c1230 Ancrene Riwle (Corpus Camb.) (1962) 67 Pellican..huau0 an oðer cunde. þet hit is aa leane. for þi..duau0 euene0 him þer to in ancre persone.
8 The mis- prefix can also be found in Romance borrowings, via the French mes-prefix.
9 Other types of Germanic etymologies have been proposed for words in AW. Zettersten (1965) identified words, possibly of Middle Low German or Flemish origin, including snateres (n. pl.) ‘snatterings, babblings’, which is “not found elsewhere in English literature” and he assumes to be and adoption of Dutch snateren or Low German snatten (237). I did not, however, encounter words of Dutch or Low German origin in my dataset. These words were rarely abbreviated, perhaps because they belonged to a mainly oral repertoire.
10 I base this view on my experience of encoding abbreviations in TEI XML (cf. Honkapohja 2013a, 2013b). For technical reasons, it is better to treat superscripts as an open category, because in addition to a set of fairly standard abbreviations, one also encounters occasional innovative ones. If one encodes superscripts as a closed category, one ends up with some abbreviated words with only single occurrences. Combining them together provides a more manageable category. In LAEME, considering each superscript as a separate type would increase the number of abbreviation types from sixteen to nearly twice as many.
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

Table 2. The abbreviation types.

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol/description</th>
<th>Expansion/extra information</th>
</tr>
</thead>
<tbody>
<tr>
<td>conus</td>
<td>ꝯ</td>
<td>‘con-’, ‘com-’ or ‘cum-’ when used as a prefix or ‘-us’</td>
</tr>
<tr>
<td>crossrk</td>
<td>a short bar through the descender of r (ꝶ) or k</td>
<td>truncation</td>
</tr>
<tr>
<td>ct</td>
<td>ꝯ</td>
<td>‘þæt’</td>
</tr>
<tr>
<td>hook</td>
<td>ꝯ</td>
<td>‘er’ or ‘re’</td>
</tr>
<tr>
<td>hookpunct</td>
<td>A specific graphetic variant of the littera that looks like yogh or z,</td>
<td>Latin syllable ‘-us’, but typically standing for the morpheme ‘-bus’ rather than singular ‘-us’</td>
</tr>
<tr>
<td>latinet</td>
<td>&amp;</td>
<td>used as a combining form for the string ‘et’ rather than for the conjunction &amp; in cleoara</td>
</tr>
<tr>
<td>mac</td>
<td></td>
<td>a general sign of abbreviation or indicating the omission of a nasal</td>
</tr>
<tr>
<td>ns</td>
<td>Nomina Sacra</td>
<td>a closed set of abbreviations for Christian names originally taken from Greek or Hebrew. In AW mainly ihū ‘Jesus’</td>
</tr>
<tr>
<td>per</td>
<td>ꝯ</td>
<td>‘par’ or ‘per’</td>
</tr>
<tr>
<td>pro</td>
<td>ꝯ</td>
<td>‘pro’</td>
</tr>
<tr>
<td>punc</td>
<td>ꝯ</td>
<td>a general sign of abbreviation, indicating that the word is abbreviated rather abbreviating a particular string of characters.</td>
</tr>
<tr>
<td>quod</td>
<td>ꝯ</td>
<td>the past tense of OE cweþan SPEAK</td>
</tr>
<tr>
<td>sup</td>
<td>superscript</td>
<td>originally a Latin practice in which part of the abbreviated word is written above the line, often indicating a contraction</td>
</tr>
<tr>
<td>tiro&amp;</td>
<td>ꝯ</td>
<td>‘and’</td>
</tr>
<tr>
<td>trunc</td>
<td>truncation</td>
<td>an umbrella term to refer to two instances in which a word was abbreviated without an abbreviation character</td>
</tr>
<tr>
<td>yurh</td>
<td>ꝯ</td>
<td>‘through’</td>
</tr>
</tbody>
</table>

4. Results and analysis

This section presents the results. I start with descriptive statistics on the frequencies of abbreviated words and move via lists of abbreviated words to the results of linear regression, using token count of abbreviation characters as the dependent variable and abbreviation types and word etymology as the independent variables.
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thensg and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

Figure 1. Abbreviation densities per specimen.

![Abbreviation density graph](image)

Figure 1 shows that the frequencies of occurrence of abbreviations are fairly even for most samples. The frequencies for three manuscripts represented by long transcriptions (*cleoara*, *titusar*, *neroar*) are reasonably close to each other. The abbreviation density in all of them is between 10–12 per cent. The *corpar* specimen, in spite of being textually close to *cleoara* and *neroar*, contains fewer abbreviations with a density below 8 per cent. The fewest abbreviations, however, are found in the textually distinct *caiusar*, in which the proportion of abbreviated words is a mere 4 % of the word count.

Different specimens taken from the same manuscript give mixed results. The scribe (*cleoara*) and the main annotator (*cleoarb*) of Cotton Cleopatra have very similar abbreviation densities. On the other hand, the two specimens taken from the Cotton Titus (*titusar* and *tituslang2*) contain very different frequencies. The highest proportion of abbreviated words, by far, is found in the mixed-dialect specimen *tituslang2*, in which more than 16 per cent of words are abbreviated.

All this leads to the question which words get abbreviated and which do not, which can best be answered by looking at word lists. Table 3 displays the ten most frequently abbreviated words in each manuscript.

Table 3. The ten most frequently abbreviated words in each manuscript. Etymologically Romance words are marked with grey background. Variant spellings are in round brackets.¹¹

<table>
<thead>
<tr>
<th></th>
<th>abbreviation density</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cleoara *</td>
</tr>
<tr>
<td>2</td>
<td>*416 ‘that’</td>
</tr>
</tbody>
</table>

¹¹ The frequencies in this table come from the words lists compiled for each MSS prior to combining abbreviation counts into the main dataset (see section 4 above).
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>eft® 70 'after'</td>
<td>þūmō(n)® 5 'woman'</td>
<td>þ(ne)® 63 'in'</td>
<td>hī 33 'him'</td>
<td>hī 82 'him'</td>
<td>eft® 5 'christ'</td>
</tr>
<tr>
<td>4</td>
<td>þ 'in' 35</td>
<td>eft® 3 'after'</td>
<td>mō 55 'man'</td>
<td>þ 30 'in'</td>
<td>þ(ne)® 59 'in'</td>
<td>ð'ce 5 'grace'</td>
</tr>
<tr>
<td>5</td>
<td>lau®d 33 'lord'</td>
<td>sūne 3 'sin'</td>
<td>hī 51 'him'</td>
<td>lau®d 30 'lord'</td>
<td>hū® 51 'them'</td>
<td>þ'sun® 4 'prison'</td>
</tr>
<tr>
<td>6</td>
<td>þere®® 17 'thereafter'</td>
<td>eft®-þ ® 3 'after that'</td>
<td>lau®d 39 'lord'</td>
<td>hū® 22 'them'</td>
<td>þ® 41 'through'</td>
<td>ðhū® 4 'Jesus'</td>
</tr>
<tr>
<td>7</td>
<td>neaut® 11 'never'</td>
<td>þ(s)®science® 3 'conscience'</td>
<td>þ-to® 27 'into'</td>
<td>mō® 18 'man'</td>
<td>mōne® 41 'man'</td>
<td>þ®phē®® 3 'prophet'</td>
</tr>
<tr>
<td>8</td>
<td>ð's't 11 'christ'</td>
<td>frō® 1 'from'</td>
<td>þē® 27 'then'</td>
<td>ou® 18 'over'</td>
<td>lau®d 40 'lord'</td>
<td>þ®gatorie®® 3 'purgatory'</td>
</tr>
<tr>
<td>9</td>
<td>þ(ero-þ®® 10 'who'</td>
<td>mōnes®® 1 'man's'</td>
<td>eaut® 24 'ever'</td>
<td>hpē® 18 'when'</td>
<td>eu® 35 'ever'</td>
<td>g'm® 4 'grim'</td>
</tr>
<tr>
<td>10</td>
<td>be®e®® 9 'better'</td>
<td>nā® 1 'none'</td>
<td>spekē®® 23 'speak'</td>
<td>strēgō®® 17 'strength'</td>
<td>sū® 31 'some'</td>
<td>þ®de®® 2 'pride'</td>
</tr>
</tbody>
</table>

The table reveals that small function words dominate the frequency counts. The two most frequent abbreviations are the same in all but one specimen. The most frequent abbreviation is always *tiro&*. The second most frequent abbreviation is in all but one specimen the *ct*. The number of tokens for these two is so high that, for example, in *cleoara* the *tiro&* and *ct* together (1,130 tokens) constitute 69.4% of all the abbreviations in the manuscript.

The second observation to be made is how many of the most frequently abbreviated words are Germanic. Very few Romance loan words make it to the list. *Cleoara* has *c's't* ‘christ’ as number eight with 11 tokens. *Cleoarb* has *conscience* with 3 tokens, *corpar* has *ðhū® 'Jesus’* as number four with 18 tokens. The scarcity of Romance vocabulary in the list of most frequently abbreviated words can partly be explained by the fact that much of the function vocabulary in English remained Germanic; however, also many lexical words in the list are Germanic, including *lau®d® 'lord’, mō® 'man’, þūmō® 'woman’, and the comparative be®e® 'better’ – which are all found in more than one specimen.

There is, however, one notable exception. The scribe of the textually distinct *caiusar* specimen does not use the *ct* or abbreviate any function words apart from ‘and’. Consequently, his list consists mainly of content words. What is more, all but three of the words appearing on the list are etymologically Romance: *c's't* ‘christ’, *ð's'ce® 'grace’, *þ'sun® 'prison’, *ðhū® 'Jesus’, *þ®phē® 'prophet’, *þ®gatorie® 'purgatory’ and *þ®de® 'pride’. Overall, as Figure 2 shows, in the two manuscripts which contain the lowest percentage of abbreviations (*caiusar* and *corpar*), the proportion of Romance words in the population of abbreviated words is higher.

Figure 2. Etymologically Romance words.12

---

12 Figure 2 omits the two most common abbreviations, Tironian & and crossed-thorn, but even when they are excluded, the percentage of Romance vocabulary amounts to no more than 20 per cent in most manuscripts.
4.1 The frequencies and distribution of abbreviation types

In this section I examine how the different types of abbreviations are distributed. Figure 3 shows the overall frequency of all abbreviation tokens in the corpus.

Figure 3. Abbreviation type/token ratio.

The graph shows that just four abbreviation types are used for the vast majority of abbreviation in all the specimens. Altogether the most frequent, *tiro&, ct, mac* and *hook*, represent 92.1 % of abbreviation in all of the manuscripts.

Statistical tests indicate that the results are highly statistically significant for both the distribution of abbreviation types and their use in different languages. Running the Pearson chi
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

square test normality yields a p-value = 4.324e-10 and the Shapiro-Francia test a p-value = 3.559e-05 (I used the functions provided by the nortest package in R). The resulting distribution, however, is not the normal distribution, or ‘bell curve’ commonly found in normally distributed data in social sciences. It also differs from the distribution predicted by Zipf’s Law which one might expect to find in language data. As Figure 4 shows, the distribution of the five most frequent abbreviation types differs from the pattern predicted for complex systems.

Figure 4. Zipf. The dark grey line shows the distribution predicted by Zipf’s Law, the light grey line is a regression line showing the best ‘fit’.

Because the distribution is non-normal, a non-parametric significance test was employed (a similar approach is used by Nurmi et al. 2018: 182–195). The Wilcoxon rank/sum test gave a p-value of 0.01703 (W = 2932).

Figure 5. Scatterplot of the abbreviation types.

---

13 While the mean is 470.6875, the median is only 43 and the standard deviation as high as 899.607301, indicating that the distribution is anything but normal.
Variation between different manuscripts is large, which is illustrated by Figure 5. It displays the same data, giving a fuller description of the range of variation, using the visreg (visual recognition) package in R. The dark line shows the mean frequency for each abbreviation type across all seven specimens. The y axis shows the number of tokens, the lighter grey band shows the confidence interval for expected values\(^{14}\) and the individual dots are the residuals, i.e. the frequencies for each abbreviation type in each specimen.

As is apparent from the figure, the values of most residuals are off the confidence interval, but they serve very well to illustrate the range of variation one can find in the frequencies of abbreviation tokens in these specimens. For example, the frequency of \textit{tiro\&} ranges between 714 and 46 across different specimens. The scatterplot of residuals is all outside of the confidence interval for all of the most frequent abbreviation types (\textit{Tiro\&}, \textit{ct}, \textit{mac} and \textit{hook}).

4.2 Abbreviation types in Germanic and Romance vocabulary

Looking at the scatterplots for etymologically Germanic and etymologically Romance words reveals interesting variation. Figure 6 shows the visreg plot for the ten most frequent abbreviation types in the Germanic vocabulary.\(^{15}\)

---

\(^{14}\) In practice, the grey band is biggest for abbreviations which only occur in one manuscript, and slightly wider for \textit{punc} which occurs in some.

\(^{15}\) The model is statistically significant with:

<table>
<thead>
<tr>
<th>Residuals:</th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>-428.71</td>
<td>-6.71</td>
<td>0.00</td>
<td>8.00</td>
<td>364.43</td>
<td></td>
</tr>
</tbody>
</table>

Residual standard error: 125.3 on 53 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared: 0.6439, Adjusted R-squared: 0.5498
F-statistic: 6.845 on 14 and 53 DF, p-value: 1.06e-07
The distribution of words tagged as Germanic has similarities to the combined distribution shown in figure 3. Thus, *tiro&* and *ct* are the two most frequent types, followed by *mac* and *hook*. *Sup*, *conus* and *per* are used with very low frequencies. *Yurh* is only used in one specimen, but is reasonably frequent in it with 50 tokens. Two abbreviation types *ns* and *pro* have a frequency of 0. The scatter plot also reveals the great amount of variation between individual specimens.

Figure 7. Scatterplot of abbreviation types for etymologically Romance words.

Figure 7, on the other hand, displays the ten most frequently abbreviated Romance abbreviation types and reveals a very different pattern. Here the two all-purpose abbreviations *hook* and *mac*

---

16 The model is statistically significant with Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals:</td>
<td>-24.286</td>
<td>-1.714</td>
<td>0.000</td>
<td>1.000</td>
<td>22.714</td>
</tr>
</tbody>
</table>

Residual standard error: 8.803 on 53 degrees of freedom
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

are also frequent, but the most frequent type are superscript abbreviations. In addition, the abbreviation types *conus*, *ns*, *pro* and *per* are used reasonably frequently. The two highest frequency items, *tiro* and *ct*, however, are used exclusively with Germanic vocabulary.

When it comes to the spread of abbreviation types across etymologically Romance and Germanic languages, the picture is a very clear one. Different abbreviation types occur with Germanic or Romance lexic gist overwhelmingly, and often exclusively. Six abbreviation types appear exclusively in Germanic words: *tiro*, *et*, *yurh*, *latinet*, *quod* and *trunc*. Three are exclusively used in Romance words: *pro*, *ns* and *hookpunct*. The remaining seven are used for both, but the distributions are far from even. As Figure 8 shows, 90 per cent of words abbreviated by *hook* or *mac* are Germanic, and *per* and *conus* are used slightly more in Germanic. *Punc* and *sup* are used more commonly in Romance words, but here one has to keep in mind that *crossrk* and *punc* are both very rare in the material, with low token counts.

Figure 8. The proportion of etymologically Germanic and Romance words appearing with abbreviation types that can be used for both.

One pair of abbreviations is particularly interesting, namely, *pro* and *per*. The former is exclusively Romance, while the latter is used for both etymological categories. Both belong to the family of p-abbreviations, which emerged during the early Middle Ages, possibly as shorthand (Hasenohr 2002: 80). Both were applied to French easily and frequently (Hector 1958: 36). However, with the AW data their distribution is clear: *per* is used for both Germanic and Romance, *pro* is Romance exclusively (see Table 4).

Table 4. The frequencies of *pro* and *per* used in etymologically Germanic and Romance words.\(^\text{17}\)

\[\text{Multiple R-squared: 0.5189, Adjusted R-squared: 0.3827} \]
\[\text{F-statistic: 3.81 on 15 and 53 Df, p-value: 0.0001503} \]

\(^{17}\) The results are statistically highly significant with a p-value lower than 0.0001 (chi-square test).
The reason for the different distribution is, naturally, that the sequence ‘pro’ is not found in etymologically native words, whereas the sequence ‘per’ or ‘par’ is natural to native vocabulary (partly overlapping with the domain of hook). Thus per is found, not only, in Romance loans such as despaunce ‘desperation/despair’ or plures ‘parlours’, but also in Germanic words such as spke ‘spark’ or spe ‘spear’. The abbreviation pro, in contrast, is limited to Romance borrowings such as pessun ‘profession’ and pphē ‘prophet’. A result like this shows that scribes abbreviate syllables or clusters of letters rather than ‘English’ words or ‘French’ words.

Even though grouping words based on etymology leads to significant results, it has to be remembered that Romance and Germanic are our categories, not theirs, and the scribes do not seem to have decided which words to abbreviate based on whether they were Latin or English, but rather abbreviated certain sequences of letters. This is also shown by the fact that superscripts, which are predominantly used for Romance vocabulary, can equally well be used for Germanic words such as gm ‘grim’.

4.3 Geographical distribution of certain abbreviations

In addition to the statistical analysis, I used the mapping tool in LAEME to check the distribution of abbreviated word forms while compiling the dataset. This led to the discovery of a usage of the conus abbreviation in particular types of words with a clear geographical distribution.

Two manuscripts of the AW, Cotton Cleopatra (including both specimen cleoara and cleorarb) and Corpus Christi use the abbreviation in three short words (h9 ‘thus’, h9 ‘house’ and sp9 ‘spouse’), which belong to different word classes and have different etymologies: ‘Thus’ is Germanic function word, ‘house’ part of the core vocabulary found in all Germanic languages and ‘spouse’ a recent French loan.18

The unabbreviated equivalents of these forms are well attested in several survey points across the country. However, the short 9-forms occur in the West Midlands, at the border of Lancashire and Cheshire and in south Shropshire – and are also found in a single scribal text localised to York. Figures 9 to 11 illustrate the distribution of the short 9-words.

Figure 9. The geographical spread h9 ‘house’ abbreviation (filled dot), including compounds such as gangeh9 and anker-h9 ‘anchorite house’, and unabbreviated ‘hus’-spellings (empty dot).

---

18 According to the OED partly from “Anglo-Norman and Old French espou, espous, espus (masculine) husband, bridegroom” and partly “Anglo-Norman spouse, Anglo-Norman and Old French spuse (feminine) (10th cent.).” The word is first attested in the twelfth century.
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

Figure 10. The geographical spread *spGeorgia* ‘spouse’ abbreviation (filled triangle), and unabbreviated ‘spus’-spellings (empty triangle).
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. Current Explorations in Middle English. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

Figure 11. The geographical spread þþ ‘thus’ abbreviation (filled square), and unabbreviated þus’-spellings (empty square).
Many of the other dots in the map, not belonging to specimens of AW, refer to vernacular works closely associated with the AW, belonging either to the so-called “Katherine Group” or the “Wooing Group” (cf. Millett and Dance 2006: ix–xi).

The geographical clustering of the short ȝ-abbreviations also overlaps with the geographical clustering of one further thorn-based abbreviation: the yurh. This use of this abbreviation is highly restricted, and is, as Laing remarks, found in a rather tightly localisable group of texts, consisting of the neroar copy of AW, the manuscript Cotton Caligula A ix and texts copied by the tremulous hand of Worcester. The tight geographical clustering becomes apparent when creating a dot map, as illustrated by figure 12.

Figure 12. The geographical spread of the ȝ ‘through’ abbreviation (filled white dot)
Both the short ȝ-forms and the yurh have connections with other well-known instances of literary activity in the West Midlands: the so-called AB language (cf. Black 1999 and Smith 2000), and also with the Tremulous Hand of Worcester. Table 5 lists all of the LAEME specimens in which each of these abbreviations can be found.

Table 5. The distribution of short-ȝ and yurh abbreviations in LAEME.

<table>
<thead>
<tr>
<th>Location</th>
<th>Text (with LAEME ID)</th>
<th>ȝ9</th>
<th>spȝ</th>
<th>ȝȝ</th>
<th>ȝ (yurh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. Cheshire</td>
<td>#118 AW titusar</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>NE Cheshire</td>
<td>#122 De Wohunge of Ure Lauerd</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>York</td>
<td>#296 Cursor mundi</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE Salop</td>
<td>#260 Sawles Warde, etc</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. Current Explorations in Middle English. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

<table>
<thead>
<tr>
<th>Manuscript</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>#261 Sawles Warde, etc</td>
<td>X X</td>
</tr>
<tr>
<td>#262 St Margaret, St Juliana</td>
<td>X X</td>
</tr>
<tr>
<td>#1000 Katherine Group</td>
<td>X X</td>
</tr>
<tr>
<td>Ludlow, Salop</td>
<td>#272 AW corpar</td>
</tr>
<tr>
<td>Worcester</td>
<td>#171 Tremulous hand</td>
</tr>
<tr>
<td></td>
<td>#172 Tremulous hand</td>
</tr>
<tr>
<td></td>
<td>#173 Tremulous hand</td>
</tr>
<tr>
<td>W Worcs</td>
<td>#245 AW neroar</td>
</tr>
<tr>
<td></td>
<td>#1800 On God Ureison, etc.</td>
</tr>
<tr>
<td>unlocalised</td>
<td>#119 AW tituslang2</td>
</tr>
<tr>
<td></td>
<td>#120 Sawles Warde</td>
</tr>
<tr>
<td></td>
<td>#121 Hali Meidhad</td>
</tr>
<tr>
<td></td>
<td>#123 St Katherine, same MS as #118 titusar and #119 tituslang2</td>
</tr>
</tbody>
</table>

The table reveals that the distribution partly results from the abbreviated forms being present in the same manuscripts. Essentially, the *yurh*, which is used by the Tremulous Hand of Worcester, is also found in the stints of three scribes (#260, #261, #262) who copied works belonging to the Katherine Group to London, British Library, Royal 17 A xxvii, entry 1. These same stints also contain the *h9*-form, which is a point of overlap in the use of these abbreviations in scribal repertoires. The *yurh*-abbreviation is also used by the Tremulous Hand of Worcester and the scribes contributing to Cotton Nero A. xiv, (containing *neroar* as well as texts that constitute LAEME specimen #1800) – but they don’t use the short *g*-forms.

Other short *g*-forms are found in two copies of the AW and related texts, localisable to the West Midlands, and also in the mixed-language Cotton Titus, whose dialect is listed in LAEME as unlocalised.

Two of the forms, *h9* and *g9*, however, also appear in one specimen, which is something of an outlier: the copy of *Cursor Mundi* found in Edinburgh, Royal College of Physicians, MS (entry 3, LAEME specimen #296), which according to Laing, “has been fitted – though the localisation is very tentative” (LAEME index of sources, #296). She places the manuscript to York, noting that the “usage of all three hands in this manuscript is strongly northern” (ibid.), and that the manuscript contains some rare spellings such as ‘sli’ for SUCH, which is “not recorded […] in any other Middle English hands, and cannot well be used therefore as a feature for fitting” (ibid.). Unlike the *sli*-spelling, which is used by all three hands, the short *g*-forms are only used by one scribe (scribe C, responsible for #296).

It is thus not completely clear, whether the geographical distribution of the short *g*-forms should be extended to York, or whether they might be carried over from an earlier exemplar, possibly written in the West Midlands. At any rate, the short *g*-forms appear to be a feature of vernacular text production, connected to the unusually active vernacular text production in the area in the thirteenth century.

5. Conclusions and future research

19 Similarities between the dialect and orthography of Nero A. xiv and the Tremulous Hand were noted by Franzen (2003), which, according to Millett and Dance, suggests “the coexistence of both traditions in the same milieu” (2006: xxviii).
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenoos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English. Studies in English Medieval Language and Literature*, vol. 56. Berlin: Peter Lang.

This short pilot study aimed to test the use ‘Germanic’ and ‘Romance’ as diagnostic categories for distribution of abbreviations in a quantitative corpus analysis, and also the suitability of the *LAEME* for the study of abbreviations. I proposed to examine three research questions: (1) what is the proportion of abbreviated words of the full word count, (2) which words get abbreviated more frequently: Romance or Germanic, and (3) do certain abbreviations occur more frequently with etymologically Romance or Germanic vocabulary?

The results revealed that the overall density of abbreviations in the relatively small-in-size vernacular devotional books of the AW family, ranged between 4 and 16 per cent, but with four out of seven specimens, containing a frequency between 10 and 12 per cent. While abbreviations were a feature taken over from Latin writing systems, in the present data they were overwhelmingly more frequent in etymologically Germanic vocabulary. This can be mainly attributed to the abbreviation of high-frequency function words, especially ‘and’ and ‘that’ (see Table 3). When function words do not get abbreviated, the proportion of Romance abbreviations is correspondingly higher (see Figure 2).

Looking at the data in more detail revealed a number of interesting observations. One specimen *caiusar* clearly stands out as outlier, due to the much higher number of Romance abbreviations and lack of abbreviations for short Germanic function-words. Interestingly, the deviant abbreviation profile is found in a copy of the treatise that also differs textually and is addressed to ‘brethren’ rather than ‘sisters’.

The scribal hand of *caiusar* has been subject to some scholarly debate. According to the description in *LAEME*, it is “An idiosyncratic mixed book hand with cursive features influenced by contemporary document hands. The scribe was evidently writing at high speed” (*LAEME*: Index of Sources, *caiusar* #276). N. R. Ker proposed that the scribe was trained abroad due to his orthography and unfamiliarity with some English letter-shapes, such as turning a short r into a long r by the addition of a separate stroke to the descender (cf. Wilson 1954: xii–xiii, Dobson 1976: 295 and Millett 2005: xvii). However, Laing argues against this (*LAEME*: Index of Sources, *caiusar* #276), pointing out that the addition of descenders is also found in other contemporary English cursive hands (citing personal communication with Parkes, 2002) and that the scribe makes use of ‘litteral substitution sets’ similar to other Middle English scribes at the time (see, e.g., Laing and Lass 2005, 2009). Her interpretation is that “Any ‘uncertainty’ in the G scribe’s use of the Old English letter shapes” is likely to have been triggered by different litteral substitution sets in the scribe’s idiolect and perhaps the exemplar being in several different writing systems (*LAEME*: Index of Sources). Nevertheless, the results of the present study lend some support to Ker’s view that the scribe may have been unfamiliar with English writing conventions, especially if the assumption is that he was writing under a very strict time constraint.

The abbreviation and suspension system developed “to save time and space” (Petti 1977: 22). If the scribe was working quickly and incorporating cursive features, wouldn’t he be likely to abbreviate more frequently? Thaisen (2011: 80) found that Adam Pinkhurst, the scribe of two *Canterbury Tales* manuscripts, used more short forms in tales, which he seems to have completed in hurry. The *caiusar* scribe does the opposite: the abbreviation density of the *caiusar* scribe is, by far, the lowest (see Figure 1, above). Moreover, when writing cursively, certain words are generally more likely to be abbreviated than others. For example, according to Hasenohr (2002) the words that get abbreviated are ones that ‘come often under the pen’, that is, short function words, endings and forms of the verb to be.20 Yet, it is precisely these

---

20 La plupart de ces abréviations ont été créées aux premiers siècles de notre ère pour être utilisées dans l’écriture courante; elles affectent surtout les désinences, les adverbes, particules et pronoms (relatifs, démonstratifs), ainsi que les formes du verbe être, qui reviennent souvent sous la plume (Hasenohr 2002: 80). [Most of these abbreviations were created in the early centuries of our era for use in cursive writing; they mainly affect the endings, the adverbs, particles and pronouns (relative, demonstrative), as
This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. Current Explorations in Middle English. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.

abbreviations that are missing. The caiusar copy stands out from the rest, because the scribe abbreviates fewer short function words (see Table 3 above). More specifically, he never uses the thorn-based abbreviations that are applied by all others. This seems to imply either unfamiliarity or, perhaps, intentional avoidance of these very common English types of abbreviations.

More generally, this result suggests that the abbreviation practices of treatises copied for different institutions, such as female or male religious houses, can vary and that the present methodology can bring to light differences that have not previously been noted. The majority of early copies of AW are connected to West-Midland literary centres of activity. These centres have received attention due to their distinctive orthographic and punctuation practices. The present study uncovered what I have called the short-9 forms of abbreviation, which are used almost exclusively in a handful of West-Midland counties. These results together show that there is interesting variation to be found using approaches that look at the abbreviations in detail.

When it comes to the main research question investigated by this study, how well do ‘Romance’ and ‘Germanic’ work as diagnostic categories, the answer seems to be that they work rather well, but with some reservations – such as the long time period from which Latin loans date. The enquiries revealed clear divisions. Nine out of sixteen, the majority, of the abbreviation types under investigation were used exclusively for either Romance or Germanic vocabulary. Seven could be used for either, but most of these showed a strong preference for one or the other. From the point of view of using these as categories in a larger-scale statistical enquiry, the results are encouraging.

Much remains to be done. One area of interest would be examining, when and where do the novel forms of abbreviation for certain high-frequency content words take over. Good candidates would be both the Tironian ȝ and the ȝ abbreviation for ‘that’, which have the highest frequency in present data. Both were eventually replaced by Romance forms: the Tironian character by the ampersand & and the crossed thorn by a superscript þ ‘that’. The present study has revealed that there is variation to be found and a properly constructed corpus resource allows us to study it productively.

References


[1] I am grateful to Dr. Raffaela Baechler for help with the translation from French.

[21] The only exceptions are occasionally using the &Tiro for ‘and’ and a using the mac for þ ‘in’ (1 token) and stû ‘some’ (1 token).
This is the author's original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.


This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.


This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.


This is the author’s original version (AOV). This article has been accepted for publication in the volume Merja Stenroos, Martti Mäkinen, Kjetil V. Thengs and Oliver M. Traxel (eds.). 2019. *Current Explorations in Middle English*. Studies in English Medieval Language and Literature, vol. 56. Berlin: Peter Lang.


