



Quantifying Medieval Scribal Habits

The Case of Abbreviations in Old West Norse Manuscripts

By Lasse Mårtensson, Eva Pettersson and Veturlíði Óskarsson

It has often been pointed out that use of abbreviations is characteristic of Old West Norse manuscripts, especially the Icelandic ones. This article deals with the use of abbreviations primarily in Old Icelandic manuscripts in comparison to Old Norwegian manuscripts and also one Old Swedish manuscript. The material consists of digital editions of manuscripts or parts of manuscripts, mostly from the digital archives of Menota and Emroon. The Icelandic manuscripts in this study are from around 1280 to the beginning of the 16th century, while the Norwegian ones are from 1200–1350, mostly from 1270–1325. The Old Swedish manuscript is dated to around 1280. Among the Icelandic manuscripts, a gradual development can be seen over time towards more abbreviation; in the manuscript from 1280, 33% of the words are abbreviated, while in the manuscript from the beginning of the 16th century, 62% of the words are abbreviated. Such a development is not observed in the Norwegian manuscripts, which were, however, written over a shorter period of time. Another tendency is that the degree of abbreviation is lower in poetry, where there is a high proportion of unusual words compared to ordinary prose. A significant regional variation in the use of abbreviations can also be observed. The Icelandic manuscripts are considerably more abbreviated than the Norwegian ones; the former vary between 33–62% abbreviated words, while the latter between 10–26% abbreviated words. Furthermore, the Norwegian manuscripts are more abbreviated than the Old Swedish manuscript examined in this study. This manuscript has 5.6% abbreviated words, which is a significantly lower percentage than in the Norwegian manuscripts.

1 Introduction

Abbreviations were an integral part of the medieval script with Latin letters, as a way of increasing writing speed and saving parchment.¹ Originating in Latin texts, the system was transferred into texts in vernacular languages, although to a varying degree. In the Nordic medieval vernaculars, the West Norse area (Norway and Iceland) made greater use of abbreviations than the East Norse area (Sweden and Denmark). Furthermore, within the West Norse area, the Icelandic manuscripts are especially characterized by an elaborated use of abbreviations. In the *Menota handbook*, ch. 6.1, it is stated that as much as one-third of the words in some Icelandic manuscripts are abbreviated.

The present study focuses on the Old Icelandic abbreviation system, but we conclude by comparing the Old Icelandic abbreviation system with the Old Norwegian one.² Our aim is to extract as much information as possible regarding the use of the abbreviations from digital editions of Old Icelandic texts, by combining digital methods for information extraction and qualitative analysis of the retrieved data. We have used digital editions of manuscripts that are in a format containing information about which letter sequences are abbreviated. For more on the material, see Section 3 below.

The study answers four separate, but interrelated, questions:

1. Does *the frequency with which abbreviations are used* increase or decrease during the Middle Ages in Old Icelandic manuscripts, and if so, to what degree?
2. What are *the most common abbreviated sequences* in the Old Icelandic scripts? Are these associated with specific high-frequency word forms (i.e., lexical abbreviations), or are they graphemic se-

1. This investigation has been carried out as part of the project *Digitization of the West Norse Manuscripts in Swedish Collections*, funded by the Swedish Riksbankens jubileumsfond (IN21-0003). We would like to thank Odd Einar Haugen, Alexandra Petrulevich and Ludvig Rödland for valuable comments. Also, we thank Per-Axel Wiktorsson for letting us use his digital edition of Cod. Holm. B 59 (Äldre Västgötalagen).
2. We have also included the only Old Swedish manuscript from the same time period as the Norwegian ones that was available to us in the required format, namely Cod. Holm. B 59 (Äldre Västgötalagen). A more in-depth investigation of the East Norse abbreviation system is a topic for future research.

- quences existing in any word (i.e., graphemic/phonemic abbreviations)?
3. Are there any common features of the words that are abbreviated and not abbreviated, respectively, regarding frequency or linguistic characteristics?
 4. What differences with respect to the use of abbreviations can be observed between Old Icelandic and Old Norwegian?

Indeed, for all the questions above, tentative answers, or at least rather rough estimations, can be given beforehand. It has been stated that the very earliest Icelandic manuscripts are relatively modestly abbreviated but that the use of abbreviations increases when we reach the 13th century (Hreinn Benediktsson 1965: 86). A hypothesis could be that this tendency is further strengthened over time during the Middle Ages (nr. 1). Also, one could assume that the sequences being abbreviated would to a large extent be associated with highly frequent word forms, since there is a very strong tendency for highly frequent words to be abbreviated (nr. 2). Furthermore, it is very likely that word frequency affects the proneness of the scribe to abbreviate (nr. 3). Finally, it is often pointed out that the tendency to abbreviate is stronger in Iceland than in Norway (nr. 4). However, our empirical basis and our digital methodology will allow us to quantify the use of abbreviations in West Norse manuscripts in a way that has not been done before.³ Even though only a limited set of manuscripts are examined, the material still covers the time periods and the geographical areas that we wish to contrast. Of course, when more material becomes available, the picture given by our investigation will be further nuanced.

We have chosen to quantify the degree of abbreviation (question nr. 1) in our investigated texts on the basis of the proportion of words being abbreviated in some way, i.e. the proportion of words that contain one or several abbreviations. In this measure, it does not matter if a word contains one or several abbreviations. Thus, a word containing one abbreviation and a word containing three are not differentiated. Both count

3. When this article was completed, the article *Abbreviation in Old Norse Manuscripts* by Tarrin Wills (2024) had been published. He addresses very similar questions as we do in our article, and uses partly the same texts for quantifications. At this late stage, we could not take Wills' investigation into account, but we can conclude that his conclusions to a large extent align with ours.

as abbreviated, as opposed to a word that contains no abbreviations. We calculate this measure on the basis of the texts in their entirety.

However, we have also made a complementary quantification of the degree of abbreviation. This is done through measuring the number of occurring abbreviations in 1,000 words. This measure thus does not state the amount of words having been abbreviated in some way (as in the former measure). Instead, it states the frequency with which abbreviations occur generally, also taking into account cases such as e.g. when two abbreviations occur in one word.

Previous research has investigated the use of abbreviations in specific manuscripts, e.g. Lindblad (1954), de Leeuw van Weenen (1993), Johansson (1997), Mårtensson (2011) and Kjeldsen (2013); cf. also Haugen (2002: 830) and Karlsson (2002: 835). The amount of attention given to the abbreviations in these investigations varies, but they certainly play a role to some extent in all of them. The present investigation builds on the knowledge coming from the studies just mentioned, but the perspective is different. Here, we want to take a contrastive view of the abbreviations according to the research questions stated above, i.e. mapping variation over time and geographically, based on all available data meeting the demands of our method. As seen in the following section, the material from some time periods is limited.

2 Background

A common formal classification of the medieval abbreviations is that of dividing them into *suspensions* (the end of the word is abbreviated), *contractions* (the middle of the word is abbreviated) and *superscript letters* and *special signs* (e.g. Hreinn Benediktsson 1965: 85–91; Driscoll 2009: 14). Regarding the content of the abbreviations, i.e. what linguistic units they represent, they can be divided into two types: 1) abbreviations representing lexical or grammatical forms and 2) abbreviations representing graphemic-phonetic units (Hreinn Benediktsson 1965: 94). An example of the former is ‘G.’ for a name, *Gunnlaugr*, *Grettir* etc., a suspension where the end of the word is abbreviated with a dot. The ‘G’ together with the dot represents a lexical unit as a whole, the name *Gunnlaugr* or *Grettir* etc. (in different cases depending on the context). The dot does not represent a specific graphemic-phonemic sequence as such. On the other

hand, the so-called *ra*-abbreviation (usually having the form of a supra-linear ‘w’, i.e. [~]) in the sequence ‘fra’ for the word *frá* (‘from’) represents the specific letter sequence ‘ra’, as it can have this reference in many other word forms, such as ‘gra’ for *grá* (‘gray’) or ‘draga’ for *draga* (‘draw, drag’).

The words that are lexically abbreviated are very often highly frequent words, either names that occur particularly often within a certain text (e.g. *Gunnlaugr* and *Grettir* in *Gunnlaugs saga* and *Grettis saga*), or other words recurring in many texts (e.g. verbs like *segja* or *mála*).

Many of these highly frequent, abbreviated words belong to the *suspension* type. It can be argued that there is a difference in the type of abbreviation between abbreviating e.g. *Gunnlaugr* with ‘G’ + dot and (forms of) *segja* (‘say’) with ‘s’ + dot. The former is confined to a certain context; thus, in e.g. *Grettis saga*, ‘G’ + dot would be used for the name *Grettir*. The latter can be used irrespective of text for the same word forms, and is thus less context-dependent. Still, the suspensions are often dependent on the interpretation of the reader (and the editor). For instance, the suspension ‘s’ + dot can be used for both the present tense form and the past tense form of *segja*. A result of this is that suspensions are often treated differently by editors, as the degree of interpretation is higher than in most other abbreviation types. We have not treated them differently in our investigation, but it must be remembered that the interpretation of some abbreviations is dependent on the editor’s choice.

It should be noted that one and the same abbreviation sign can be used for both lexical and graphemic-phonetic representation; superscript ‘m’ can be used in lexical abbreviations, e.g. ‘r’ + superscript ‘m’ for *riddurum* (dat. of *riddari* ‘knight’) as well as for the sequence ‘um’, occurring in different words. A central aspect of our investigation is therefore to make a distinction between the lexical and the graphemic-phonemic types of abbreviations, as they are conditioned by different features of the lexical items.

As stated above, there is an abundance of abbreviations in the West Norse, especially the Icelandic, script. It is known that recurring formulas are often abbreviated through suspensions, e.g. in manuscripts with poetry or legal texts (e.g. Hreinn Benediktsson 1965: 87). For this reason, the genres of the texts in our investigated manuscripts have to be accounted for in the present analysis.

3 Data

The chosen methodology requires digital transcriptions of manuscripts in a format where abbreviations are indicated, i.e. diplomatic editions of these manuscripts. Furthermore, we need texts of a certain length in order to make the quantification reliable, since the use of abbreviations in shorter texts may not be representative for the manuscript as a whole. We have set the threshold to 10,000 words. We have used all available texts meeting these demands in the digital text archives Menota (“Medieval Nordic Text Archive”) and Emroon (“Etymologically and Morphologically defined Referential Orthography for Old Norse”, Robert Paulsen, Bergen), and also a digital edition in Menota format of the manuscript DG 10, not yet present in the archives.⁴ Within this text corpus, we have texts from medieval Iceland, Norway and Sweden, enabling a comparison between these countries. The Icelandic material contains manuscripts from different times, allowing for the investigation of chronological development. We are aware that one or two manuscripts from a certain century are not fully representative for the century as a whole. As stated, our selection of data has been restricted by the material available in the required format, and thus the investigated manuscripts have to function as representatives for their centuries in this study. The abbreviation system in a manuscript is not only conditioned by chronology, but also by the individual scribe. We know, for example, that a number of the Norwegian legal manuscripts were produced by the same scribe. The individual aspect, however, is not part of this study. When more data is available, the results of the present investigation can be further refined.

In total, the Icelandic dataset contains 525,158 words (i.e. tokens) from the time period 1280–1500, and the Norwegian dataset consists of 706,500 words from the time period 1200–1350, as illustrated in Tables 1 and 2 below. A large part of the Norwegian manuscripts is from a rather limited period, 1250–1325, even though there are a few earlier and later ones. Furthermore, it is not possible to investigate Norwegian manuscripts for the entire period of the Icelandic manuscripts (i.e. until the beginning of the 16th century), as very few manuscripts were produced

4. The texts in Emroon follow the Menota standard. The text in DG 10 is transcribed by one of the authors of this article, Veturlíði Óskarsson.

in Norwegian after 1350, and almost none after 1400 (Haugen 2018: 228–229; Mørck 2018: 299).⁵

In the Icelandic corpus, all manuscripts contain mainly narrative texts, although especially DG 11 and AM 242 fol (both containing the work *Snorra Edda*) include a large amount of skaldic poetry. Thus, genre is not included in Table 1. In the Norwegian corpus, on the other hand, there is a mixture of genres, as stated in the column furthest to the right (Table 2). Most of the manuscripts in the corpus are transcribed in full. For more detailed information, see the metadata about the texts on the Menota and the Emroon homepages. For Swedish, we currently only have access to one text, as illustrated in Table 3. This text, a digital edition of Cod. Holm. B 59, containing the Old Swedish law text *Äldre Västgötalagen*, was given to us by Prof. Per-Axel Wiktorsson (emeritus). It is not in Menota format, and the information regarding abbreviations was extracted on the basis of the sequences in italics in the text, signaling abbreviations. In the tables below the datings are taken from ONP.

Table 1 Old Icelandic data set used throughout our study (for the corpus column, *E* stands for Emroon, *M* for Menota, *MT* for Menota test archive and *U* for Unpublished).

Manuscript	Title	Corpus	Ms Dating	Words
AM 519 a 4to	Alexanders saga	M	1280	46,732
AM 45 fol	Codex Frisianus, misc. texts	E	1300–1325	11,980
DG 11	Codex Upsaliensis, e.g. Snorra Edda	MT	1300–1325	35,876
AM 132 fol	Möðruvallabók, misc. texts	M/MT	1330–1370	177,618
WolfAug 9 10 4to	Egils saga Skalla-Grimssonar in Wolfenbüttelbók	M	1330–1370	38,617
AM 573 4to	Trójumanna saga, Breta saga, Valvens þáttir	E	1330–1375	14,238
AM 242 fol	Codex Wormianus, e.g. Snorra Edda	M	1350	62,586
AM 764 4to	Reynistaðarbók, misc. texts	E	1360–1380	35,042
AM 557 4to	Skálholtsbók, e.g. Eiríks saga rauða	MT	1420–1450	45,081
DG 10	Grettis saga Ásmundarsonar	U	1500	57,388
			1280–1500	525,158

5. Bjørn Bandlien (2014: 251–252) points out that one of the scribes of DG 4–7, normally counted as Norwegian, may have been an Icelandic.

Table 2 Old Norwegian data set used throughout our study (for the corpus column, *E* stands for Emroon, *M* for Menota).

Manuscript	Title	Corpus	Ms Dating	Words	Genre
AM 619 4to	Old Norwegian homily book	M/E	1200–1225	62,237	homilies
DG 8 II	Óláfs saga ins helga (legendary)	M	1225–1250	41,142	narrative
DG 4–7	Elíss saga ok Rósamundar	M	1270	17,548	narrative
DG 4–7	Strengleikar	M	1270	38,453	narrative
AM 243 bǫ fol	Konungs skuggsjá	M	1275	63,910	didactic
Holm perg 6 fol	Barlaams ok Josaphats saga	M	1275	76,411	narrative
Holm perg 34 4to	Bójarlög, Farmannalög Magn. Hákon.	M	1275–1300	11,283	law
Holm perg 34 4to	Landslög Magnúss Hákonarsonar	M	1275–1300	45,243	law
AM 302 fol	Landslög Magnúss Hákonarsonar	M	1300	50,188	law
AM 305 fol	Landslög Magnúss Hákonarsonar	M	1300	47,367	law
AM 56 4to	Landslög Magnúss Hákonarsonar	M	1300	49,256	law
Holm perg 17 4to	Thómass saga erkibiskups	M	1300	59,884	narrative
AM 78 4to	Kristinréttir, Bójarlög Magn. Hákon.	M	1300	45,516	law
Holm perg 30 4to	Landslög Magnúss Hákonarsonar	M	1300–1325	47,948	law
DG 8 I	Landslög Magn. Hákon. Farmannalög	M	1300–1350	50,114	law
			1200–1350	706,500	

Table 3 Old Swedish data set used throughout the study.

Manuscript	Title	Ms Dating	Words	Genre
Cod. Holm. B 59	Äldre Västgötalagen	1280	18,634	law

It should be noted that the Icelandic material from the 14th century is larger than from the 13th, 15th and 16th centuries. The three last mentioned centuries are represented by one manuscript each, whereas there are seven manuscripts from the 14th century. This is solely determined by the data available in the Menota and Emroon archives.

4 Quantifications

In this section, we present the results for each of the research questions listed in the Introduction.

4.1 Use of abbreviations over time
Does the frequency with which abbreviations are used increase or decrease during the Middle Ages in Old Icelandic manuscripts, and if so to what degree?

To answer this question, we extract all word forms in our corpus that contain the Menota <ex>-tag, signalling the use of an abbreviation. The digital texts have been prepared by different editors, and even though all are following the mark-up system of Menota, there may be some minor differences in how the rules have been applied. However, this will not affect the general tendencies emerging from the extractions. The results are presented in Figure 1, with the oldest texts to the left, and the youngest to the right. The percentage numbers in the figure state how many words contain an abbreviation (or more than one). The remaining words are thus not abbreviated.

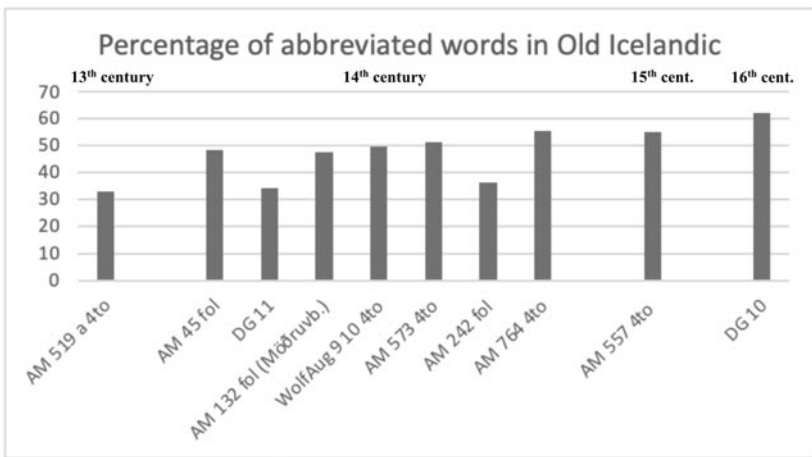


Figure 1. Percentage of abbreviated words in the Old Icelandic corpus, with the oldest manuscript to the left, and the youngest to the right.

In Figure 1, *Möðruvallabók* is displayed as one unit. We have used the eight texts that were available in Menota at the time.⁶ These eight texts show rather large similarities in abbreviation tendency (44%–51%). The variation could still indicate that the exemplar has played a part in the scribes' use of abbreviations. A further investigation of this aspect is a task for future research.

6. *Bandamanna saga*, *Droplaugarsona saga*, *Finnboga saga*, *Fóstbrœðra saga*, *Hallfreðar saga*, *Kormáks saga*, *Víga-Glúms saga* and *Ölkofra þáttr*.

Based on our text collection, the hypothesis that the proportion of abbreviated words in Old Icelandic manuscripts increases over time is confirmed. The oldest Icelandic manuscript in the corpus is AM 519 a 4to, dated to 1280 (*Alexanders saga*), which contains about 33% abbreviated words. This is a significantly lower percentage than the youngest manuscript, DG 10, dated to around or shortly after 1500 (*Grettis saga*), which contains about 62% abbreviated words. There is some variation in the 14th century texts, with for example 34% abbreviations in *Snorra Edda* (DG 11; see further below) and 48% abbreviations in the multitext manuscript *Codex Frisianus* (AM 45 fol), both dated to 1300–1325. However, when we calculate the average over all of the 14th century texts, the trend is clear, as illustrated in Figure 2. The most deviant manuscripts from the 14th century, DG 11 and AM 242 fol, are discussed further below.

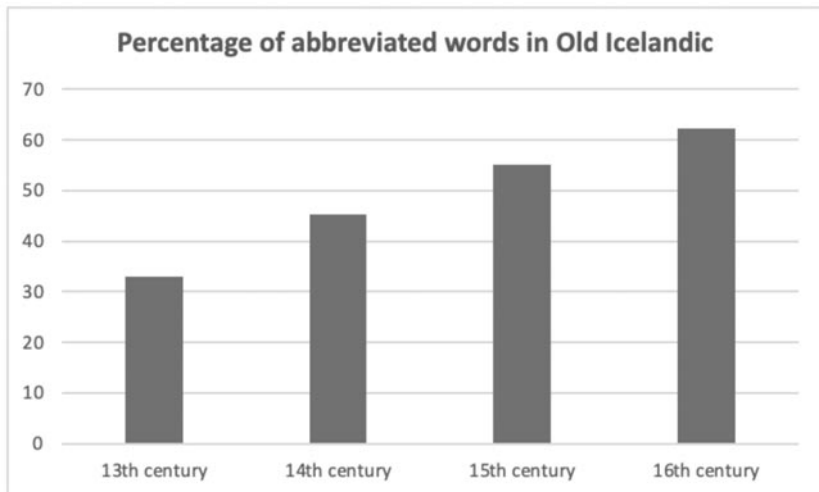


Figure 2. Percentage of abbreviated words in Old Icelandic manuscripts, divided by century.

As stated above, there is some variation between the manuscripts from the 14th century. Two manuscripts distinguish themselves in having a lower usage of abbreviations than the others: DG 11 and AM 242 fol. Both of these manuscripts contain *Snorra Edda*, with all the central parts: *Gylfaginning*, *Skáldskaparmál* and *Háttatal*. To a large extent, the content of the two latter parts is skaldic poetry, as illustrations of poetic language

(*Skáldskaparmál*) and verse meter (*Háttatal*). This type of poetry consists to a large degree of infrequent poetic words, whose meaning was probably often obscure to the medieval scribes. As touched upon in the beginning, and as will be discussed further below, the frequency of the words is an important factor for the usage of abbreviations. Thus, a high degree of infrequent, poetically marked words probably leads to a lower degree of abbreviation. The scribes may have looked upon the abbreviations as limiting the readability somewhat, thus not to be used excessively in words that are difficult to interpret in the first place.

It can finally be noted that in Section 1, we referred to the statement in the *Menota handbook* that up to a third of the words in the Old Icelandic manuscripts are abbreviated. Indeed, that number is exceeded for all manuscripts in our corpus, as seen in Figure 1.

In the investigation described above, we have measured the proportion of words that contain abbreviations. This measurement does not account for how many abbreviations are used in total, since one and the same word can contain one or several abbreviations. To investigate the total frequency with which abbreviations are used, we have calculated the number of abbreviated sequences per 1,000 words for each century. This is meant to account for instances where more than one abbreviation is used in one word. These measurements are shown in Figure 3.

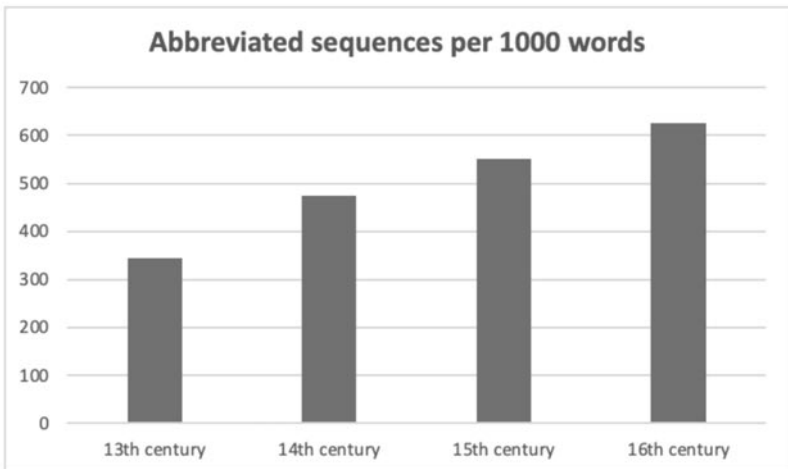


Figure 3. Number of abbreviated sequences per 1,000 words in Old Icelandic manuscripts, divided by century.

As can be seen in the figure, the trend in this measurement is the same as for the percentage of abbreviated words. There is a development towards a gradually higher degree of abbreviation, with 345 abbreviated sequences per 1,000 words in the 13th century text up to 626 abbreviated sequences per 1,000 words in the 16th century text.

The investigation in the previous section showed a distinctly lower degree of abbreviations in the manuscripts DG 11 and AM 242 fol compared to the other manuscripts originating from their respective centuries. The interpretation mentioned above was that these manuscripts were less abbreviated because they contained a large amount of poetry, and that poetry, comprising a high number of infrequent and often obscure words, was less abbreviated than prose. To investigate this aspect further, we have compared pages in the two mentioned manuscripts where the content is mainly poetry and prose, respectively. We have extracted five pages of prose and five pages of mainly poetry in the manuscripts DG 11 and AM 242 fol. As stated, the pages with poetry contain *mainly* poetry, but not exclusively, as the verse quotes are usually introduced by formulaic prose statements like “Svá kvað Einarr”. Furthermore, sections of kennings are often introduced with questions (in prose), like “Hvernig skal kenna himin?”. These sentences are usually abbreviated, maybe more so than normal prose as such introductory phrases occur very often and are predictable from the context.

The pages used for this study are listed in Table 4 below (the span given in the format used in the Menota texts, i.e. fol. in DG 11, and page in AM 242 fol). The results of the investigation are shown in Figure 4.

Table 4 Prose and poetry parts in the manuscripts DG 11 and AM 242 fol.

	DG 11	AM 242
Prose	13v–15v	p. 27–32
Poetry	27v–29v	p. 45–47, 53, 80

As can be seen in Figure 4, the assumption that poetry is less abbreviated than prose is confirmed by the investigation. The pages in DG 11 and AM 242 fol, containing mainly poetry, are less abbreviated than those containing prose. Furthermore, the mentioned introductory phrases, such as “Svá kvað X” and “Hvernig skal kenna X?”, usually abbreviated, are included in the measurement of the poetry pages. Thus, the diffe-

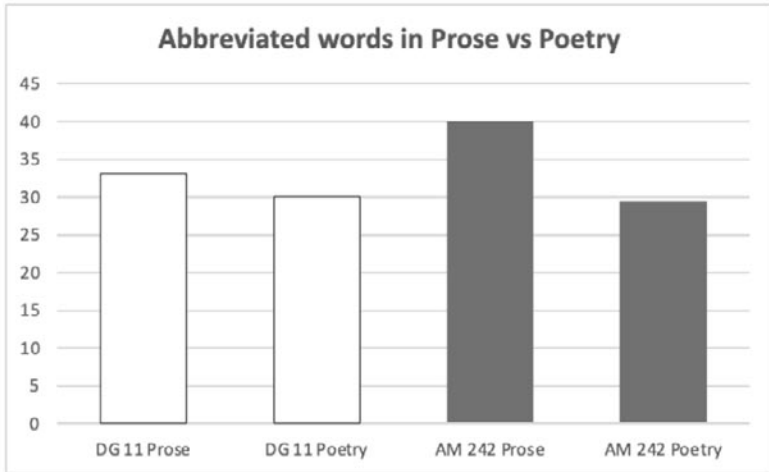


Figure 4. Proportion of abbreviated words in prose compared to poetry, in the Old Icelandic manuscripts DG 11 and AM 242 fol.

rence between poetry and prose is probably even larger than the numbers in Figure 4 suggest.

As can be seen in Figure 1, DG 11 has a slightly lower general degree of abbreviation than AM 242 fol. This could be explained within the general trend observed in the previous section, namely that there is a general increase over time in the abbreviation tendency. DG 11 was produced slightly earlier than AM 242 fol (1300–1325 vs. 1350), so one could argue that the difference has a chronological explanation. However, it is worth noting that the degree of abbreviation on the poetry pages is almost the same in the two manuscripts, and the difference is mainly present in the prose. On the basis of our investigated material (which is rather small for far-reaching conclusions in this aspect), it appears that the tendency to abbreviate poetry, i.e. infrequent and obscure words, remains stable over time, whereas the chronological change towards higher frequency of abbreviation takes place within prose, to a large extent consisting of frequent and well-known words.

4.2 Abbreviated sequences

What are the most common abbreviated sequences in the Old Icelandic scripts? Are these associated with specific high-frequency word forms (i.e., lexical abbreviations)?

viations), or are they graphemic sequences existing in any word (i.e., graphemic-phonemic abbreviations)?

As stated in the beginning, a distinction is often made between abbreviations denoting lexical units and those denoting separate graphemic-phonemic sequences. The former are usually fixed writings of highly frequent or otherwise predictable word forms, whereas the latter are abbreviations used for sequences, normally two graphemes (representing a vowel and a consonant), occurring in different words. In this section, the abbreviated sequences in the manuscripts are investigated, and initially, no difference is made between the two types. All abbreviated sequences in the manuscripts are extracted according to the description below, but in the discussion of the results, the two abbreviation types will be analyzed. The graphemic-phonemic type would be possible mainly in the cases where only two graphemes are abbreviated, and when the sequences are longer, it is most likely lexical units that are abbreviated.

To investigate the abbreviated sequences in the Old Icelandic manuscripts, we use the diplomatic version of the text to identify the character sequences that have been abbreviated, by extracting the character sequences occurring within the *<ex>* tags in the XML version of the source file. For example, if the facsimile version *ðott* and the diplomatic version is *dott <ex>er</ex>*, the character sequence *er* will be extracted as an instance in the set of abbreviated sequences for that manuscript.

Since the manuscripts are of different length, and thus not directly comparable, we extract a random set of 40,000 words for each century and calculate the total number of abbreviated sequences for this subset of the corpus, as well as the total number of abbreviation types. Also, since there is only one manuscript available for the 13th (AM 519 a 4to), 15th (AM 557 4to) and 16th century (DG 10) respectively (see Table 1), we choose only one manuscript from the 14th century as well (the largest one, namely AM 132 fol, *Möðruvallabók*).

With this approach as a basis, we calculate the total number of abbreviated sequences for each sample, contrasted with the number of unique abbreviated sequences, as illustrated in Figure 5. From the results, we can see that the number of unique abbreviated sequences increases quite proportionally to the number of total abbreviated sequences over the centuries. This indicates that the scribes' tendency to abbreviate new sequen-

ces increased parallel to the general increase in the use of abbreviations. The only exception is the 14th century sample, for which the proportion of unique abbreviated sequences is slightly less than for the other centuries.

Here, it would be interesting to study the set of abbreviation characters used for abbreviating. However, these are not specifically marked in all of the Menota files at hand, making such a study difficult to conduct automatically.

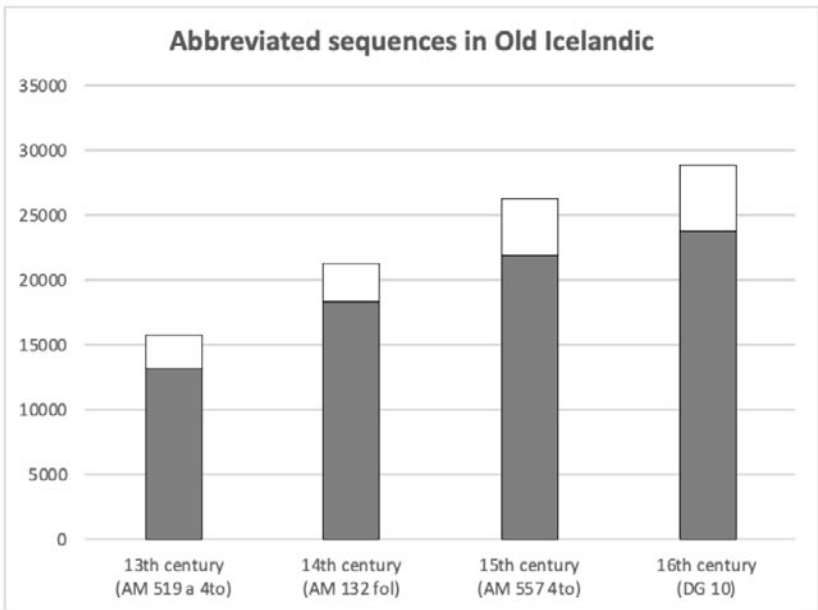


Figure 5. Total and unique abbreviated sequences in Old Icelandic manuscripts, divided by century. The white parts: abbreviated sequences occurring only once. The grey parts: abbreviated sequences occurring more than once.

When looking at the most frequent abbreviated letter sequences, it is clear that they are mainly the same over the centuries, with nasals (*n* and *m*) and frequent word endings (e.g., *er/ar*) occurring very frequently. Furthermore, most of the abbreviated sequences belong to fixed writings of highly frequent words. For instance, in all centuries except in the youngest manuscript, from around 1500 (DG 10), the sequence ‘onun’, ‘onung’ or similar is recorded among the most frequently abbreviated se-

quences. This is of course due to the fact that the word *konungr* is frequent in the texts examined from these centuries. Also sequences like 'ok'/'og' (*ok*), 'irir'/'yrir' (*fyrir*), 'ann' (*hann*), 'agdi' (*sagði*), 'igi'/'cki' (*eigi/ekki*) and 'onv' (*honum*) are used in specific words, and thus constitute lexical abbreviations. The same obviously also goes for sequences like 'lexander' (*Alexander*), 'arius' (*Darius*) and 'retter' (*Grettir*), as they are used in names, highly frequent within the texts in which they are important protagonists.

Many of the one- and two-character sequences are also mostly restricted to specific words. This is the case for e.g. 'r' (*er*), 'eð'/'ed' (*með*), 'at' (*bat*), 'il' (*til*), 'ua' (*svá*) and 'on' (*hon*). These sequences are really to be interpreted as part of lexical abbreviations rather than graphemic-phonetic abbreviations. A conclusion to be drawn from this is that the use of abbreviations is to a great extent generally lexically determined, which is to say triggered by frequent words, also when a one- or two-character sequence is being abbreviated.

However, there are also examples of actual graphemic-phonemic abbreviations, where it is clear that certain sequences are being abbreviated independently of the word in which they occur. Most clearly this is the case with abbreviated nasals, 'n' and 'm'. These elements can be abbreviated in all contexts in which they occur, and this is done with the horizontal abbreviation stroke (sometimes called *nasal stroke*). This is probably the most common abbreviation type of all, and it had a very widespread use in the Nordic countries (as well as elsewhere). For instance, in the East Norse area (Sweden and Denmark), where the use of abbreviations was more restricted than in the West Norse area, the abbreviation of nasals was very common. Also, sequences representing grammatical endings are used independently of the words. This is especially true for plural endings on nouns and adjectives, most importantly 'er'/'ir' (*-ir*; with the so-called *er*-abbreviation) and 'ar' (*-ar*; with superscript 'r'). Furthermore, the sequence 'ra' (with the so-called *ra*-abbreviation) can be included in this group. Even though it is often used in the word *frá*, it is also often used to represent the ending *-ra*, occurring e.g. in adjectives in the gen. pl. This ending, however, is not as common as *-ir* and *-ar*.

It is interesting to note the relative increase in suspensions over the years, with approximately 2.6% of the abbreviations being suspensions in the 13th century text, compared to 6.6% in the 14th century texts and

9.4% in the 15th century text. However, the youngest manuscript (DG 10) contains fewer suspensions (approximately 6.1%).

4.3 Features characterizing abbreviations

Are there any common features of the words that are abbreviated and not abbreviated, respectively, regarding frequency or linguistic characteristics?

4.3.1 Part-of-speech

To study the linguistic characteristics of words that are abbreviated and not abbreviated, respectively, we start by looking at the part-of-speech distribution. For this study, we make use of a subset of the original corpus, where part-of-speech information is available, resulting in a sub-corpus of 211,092 words in total. The manuscripts that the sub-corpus is taken from are: AM 764 4to, AM 573 4to, AM 45 fol, WolfAug 9 10 4to, AM 519 a 4to and AM 132 fol. These are all Icelandic, and all except one are from the 14th century, the exception being AM 519 a 4to, which is from the 13th century. These transcriptions all had mark-up regarding part-of-speech.

The category “Other” in Figure 6 consists of a mix of categories that would be too small to display on their own: foreign words, infinitive marker, numerals and similar. They will not be further discussed below.

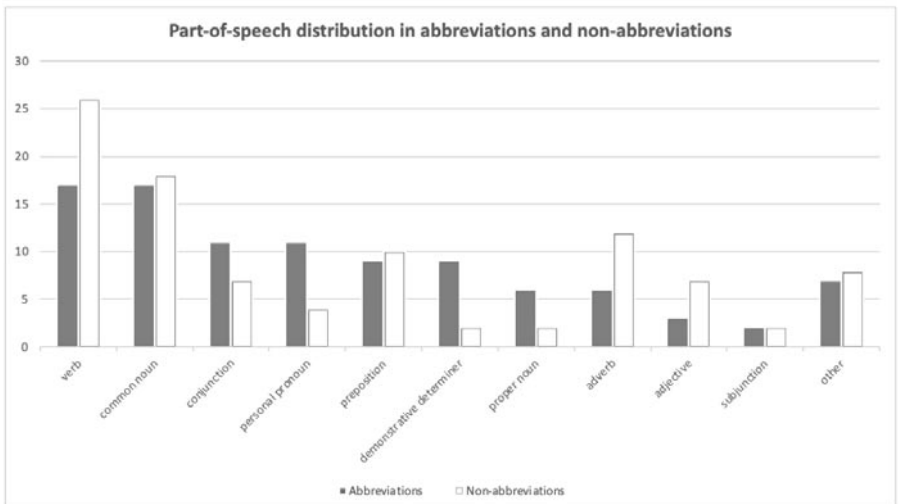


Figure 6. Part-of-speech distribution for abbreviated and non-abbreviated words respectively, in our dataset.

As seen from the statistics in Figure 6, verbs and nouns dominate among both abbreviated and non-abbreviated words, which is not surprising considering the large amounts of instances for these parts-of-speech. Interestingly and perhaps expectedly, proper nouns, demonstrative determiners and personal pronouns show a significantly higher proportion of instances in the abbreviation category.

In Figure 6, the statistics show the relative frequency for each part-of-speech, that is the number of instances divided by the total number of instances (multiplied by 100).

As stated above, some parts-of-speech are generally more abbreviated than others. As frequency is an important factor for the use of abbreviations (see discussion above), one could assume that grammatical words such as pronouns, conjunctions etc. are more heavily abbreviated than content words such as nouns or verbs. Of course, there are highly frequent words within these word classes as well; verbs like *segja* ('say') and *mála* ('say') are very frequent, and indeed they are usually abbreviated. Still, the numbers in Figure 6 suggest that a large number of verbs are written without abbreviations. Indeed, for content words like nouns, verbs and adjectives, the number of non-abbreviated writings supersede the abbreviated ones. Otherwise, the assumption that the parts-of-speech constituting grammatical words are more abbreviated than others seems to hold true. Conjunctions, pronouns and demonstrative determiners are very often written in an abbreviated way, and there are usually fixed writings for the most frequent forms: 'h' + horizontal nasal stroke for *hann*, 'þr' + horizontal nasal stroke for *þeir* etc. The sequence *þess-* in the demonstrative determiners like *þessir*, *þessar* etc. is very often abbreviated with 'þ' + horizontal nasal stroke.

Two parts-of-speech require a comment, namely prepositions and proper nouns. The prepositions, highly frequent function words, have a fairly equal distribution between abbreviated and non-abbreviated forms. This might appear surprising, but it must be remembered that two of the most frequent prepositions are *á* ('on') and *í* ('in'). They cannot be abbreviated, and thus they can be expected to significantly heighten the number of non-abbreviated forms in this class. Proper nouns, finally, are also characterized by a high degree of abbreviated writings. This is due to the previously mentioned, well-known practice of writing frequently occurring names in a certain text as suspensions, i.e. with the initial letter of the name followed by a dot (or a similar mark; cf. Guðvarður Már Gunnlaugsson 2007: 7), after having written out the name in full once.

4.3.2 Frequency

As for the frequency of the words, we have already shown (section 4.2) that words that occur frequently are often abbreviated, since these words are known to the reader and therefore do not need to be written out in their entirety each time. Here, the tendency observed in section 4.2 will be elaborated further.

In this part of the study we consider nouns only. This is to avoid having only highly frequent function words at the top of the frequency list. For the noun category, as illustrated in Figure 6 above, there is a fairly equal distribution of abbreviated and non-abbreviated forms, which justifies the investigation of the abbreviation rate for the most frequent instances of this category. In Table 5 below, the 25 most frequent noun forms in the corpus are listed, with the frequency with which each noun is abbreviated and non-abbreviated, respectively. It is interesting to note here that the overwhelming majority of the most frequent nouns are often abbreviated. For 18 out of the 25 nouns listed, a much larger proportion are abbreviated than not abbreviated. Eight of these nouns only occur in abbreviated form, and never unabbreviated. It must also be stressed here that we are primarily analyzing the *word forms* here, not the *lemma*. In the table the normalized forms are given, representing an inflected form of a noun. This is also due to the format of our material, as not all texts in our corpus are lemmatized.

Table 5 The 25 most frequent word forms.

Word form	Abbreviated	Non-abbreviated
konungr	960	0
menn	647	2
maðr	288	98
ár ⁷	272	54
manna	294	4
son	89	133
daga	172	36
fé	0	171
mánaði	154	2
konung	153	1
konungs	150	0
konungi	140	0
mønnum	133	0
ráð	0	132

7. This is the lemma *ár* ‘year’. It has been written ‘a’ + supralinear ‘r’.

mann	128	0
dag	12	115
ætt ⁸	121	0
konungs	118	0
guð	27	87
mál	0	111
hér	96	13
son	6	102
land	99	5
landi	94	0

The word forms in Table 5 represent case forms of nouns (in their normalized forms) in the Old Icelandic manuscripts, with statistics on the number of times they occur in abbreviated and non-abbreviated form.⁹

As seen in the table, several of the word forms belong to the same lemma: *konungr*, *konung*, *konungi*, *konungs* (lemma *konungr*), *maðr*, *mann*, *menn*, *manna*, *monnum* (lemma *maðr*). It should be noted that the forms of *konungr*, highly frequent, only once occur written out in full in the entire investigated corpus. This is a rather long word, and the abbreviated forms of this word saved a lot of space and labor for the scribe. The tendency is thus clear: many of the high-frequency words are often abbreviated, in some cases always. Some of the frequent words are, on the other hand, never abbreviated: *fé*, *ráð* and *mál*. This is due to the fact that the letter sequences in these words do not have an established abbreviation attached to them. For instance, the sequence ‘al’ (in *mál*) cannot be abbreviated with superscript ‘a’, and superscript ‘l’ does not occur. A self-evident but important conclusion to be drawn from this is that the structure of the words, the letter sequences constituting the words in their written form, is an important factor for the abbreviation tendency. They must contain sequences that can be abbreviated. Thus, frequency alone is not enough.

When studying frequency, it is also relevant to look at not only the most frequent instances, but also the *least* frequent ones. On this note, we observe that out of the 24,610 word forms that occur only once in our corpus, 9,235 word forms are abbreviated, and 15,375 are not. This

8. This is the lemma *ætt* ‘family, lineage’. It has been written ‘æt’ + supralinear dot to indicate long *t*.

9. We make no distinction here between palaeographic variants, like high and round ‘s’ and similar.

implies that also (very) low-frequency words are abbreviated to a rather high degree. This is probably due to the fact that nasals often are abbreviated irrespective of word type, and the same holds true for the highly frequent endings *-ir* and *-ar*.

It could be expected that longer words are more often abbreviated, as it saves both time for the writer and space on the parchment. A study of the average word length in our text collection indeed shows that the average word length of abbreviated words is 7.5 characters, while the average word length for non-abbreviated words is 6.1 characters, which strengthens this hypothesis. However, even short words are often abbreviated, like *ok*, *með* and the like.

4.4 Comparison between Old Icelandic and Old Norwegian

What differences with respect to the use of abbreviations can be observed between Old Icelandic and Old Norwegian?

As seen from Figure 7, the proportion of abbreviations in the Old Norwegian manuscripts is substantially lower than in the Old Icelandic ones. For Old Icelandic, the proportion of abbreviations varies from 33% to

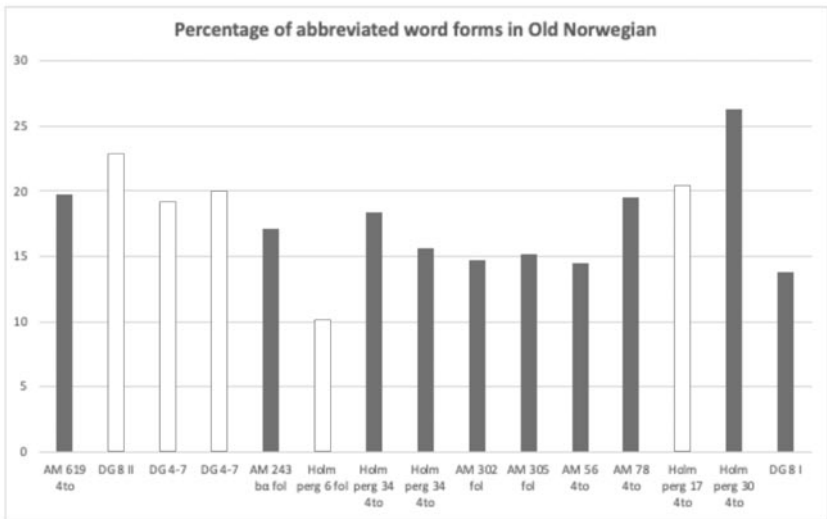


Figure 7. Percentage of abbreviated word forms in the Old Norwegian manuscripts, with the oldest manuscript to the left, and the youngest to the right (narrative texts marked in white).

62% (see Figure 1), whereas in the Old Norwegian manuscripts, the proportion varies between 10% and 26%. It should be noted that a large part of the Icelandic manuscripts are later than the investigated Norwegian ones. If we only consider the three Icelandic manuscripts from the same time period (AM 519 4to, AM 45 fol and DG 11), the proportion of abbreviations varies between 33% and 48% (Figure 1). It could be argued that the Old Norwegian manuscripts are more heterogeneous in genre, containing not only narrative texts, but also law texts, as well as religious and didactic texts. However, even if we only consider the Norwegian narrative texts (marked in orange in Figure 7), the average proportion of abbreviations is 18%, which is way below any of the Old Icelandic manuscripts.

From Figure 7 it is clear that the examined manuscripts do not display a chronological development towards a more frequent use of abbreviations, as was the case in the Icelandic material. Rather, the impression is that of variation; the manuscripts vary greatly between each other in the tendency to abbreviate. This variation can also be observed between manuscripts produced close in time. The data does not allow for very far-reaching conclusions as to the reasons for this, but a possible explanation could be that the scribes of medieval Norway display a larger degree of individual variation in this aspect, as the abbreviation system was generally more restricted in Norway than in Iceland. That is, a scribe prone to abbreviating in Norway would deviate more from the other scribes.

Regarding the Norwegian material, the same tendency as for the Icelandic material can be observed: most of the sequences are to be interpreted as lexical abbreviations, being a part of fixed writings of frequent words. Furthermore, it appears that, to a large extent, the same words that are abbreviated in the Norwegian manuscripts are also abbreviated in the Icelandic manuscripts. Highly frequent abbreviated sequences like ‘onon’ and ‘onong’ belong to the lemma *konungr*, the abbreviated sequence ‘eð’ to the lemma *með* and ‘ann’ to the form *hann*, and so on. Some cases where the abbreviation is probably to be interpreted as graphemic-phonemic are indeed present in our material, e.g. the nasals ‘n’ and ‘m’, and also in many cases the sequences ‘er’ and ‘ir’. It is notable, however, that the sequence ‘ar’, highly frequent in the Icelandic manuscripts for the ending *-ar*, is not present among the most common abbreviated sequences in the Norwegian manuscripts.

As stated earlier, we only have access to one digital text with the desired digital format in Old Swedish from the same time period as the Old Norwegian manuscripts. This is Cod. Holm. B 59, *Äldre Västgötalagen*, dated to 1280. Even though this is only one manuscript, it still makes an interesting comparison with the Old Norwegian manuscripts. It is interesting to see if there is a generally higher degree of abbreviation in the West Norse manuscripts compared to Old Swedish manuscripts, or if this tendency is only true for the Old Icelandic manuscripts. It appears, on the basis of our investigation of B 59, that there is a stronger tendency to abbreviate in the Old Norwegian manuscripts than in the Old Swedish one. B 59, i.e. *Äldre Västgötalagen*, contains only 5.6% abbreviated words, a substantially lower proportion than even the Old Norwegian manuscripts, where abbreviated words appear at a rate of 10–26%. Recall that the Old Icelandic manuscript from the 13th century, AM 519 a 4to, has 33% abbreviated words. Thus, the Old Swedish manuscript investigated here is considerably less abbreviated than the Old Norwegian ones, which in turn are much less abbreviated than the Old Icelandic ones.

Even though the abbreviation tendency is much more restricted in the Old Swedish manuscript than in the Norwegian and the Icelandic ones, the sequences being abbreviated are largely the same also in this manuscript. Abbreviation of nasals, i.e. ‘n’ and ‘m’, occurs regularly, and also the sequence ‘æp’ in the preposition *mæp* (‘with’). The sequence ‘ra’ is abbreviated in different words, e.g. ‘drap’ (‘manslaughter’) and ‘vintra’ (‘winters’). A practice that is common in Cod. Holm. B 59 is to abbreviate the word *maþar* (‘man; human’) with the **m**-rune (ʁ) having the name *maþar/maðr*. This usage is also known from the early West Norse manuscripts, but it is not used in the manuscripts used in the present investigation. As in the Norwegian and the Icelandic manuscripts, the word *konungar* (‘king’) is abbreviated in the Old Swedish manuscript with a contraction. The difference is that this is only done occasionally in the Old Swedish manuscripts, and this word also occurs written out in full. In the West Norse manuscripts, this word is almost exclusively abbreviated.

5 Conclusions and future work

In the present study, we show that several of the previously stated hypotheses about the use of abbreviations in West Norse manuscripts could be validated empirically. The use of abbreviations appears to increase gradually during the Middle Ages in Iceland. However, this tendency does not hold true for Norway. Instead, it would seem that the examined Old Norwegian material displays more incidental variation. It should be noted, however, that the time span for the Norwegian material is limited, making it hard to draw any definite conclusions on this front.

The analyses conducted above also show that the use of abbreviations was to a large extent connected with specific words and word forms. This appears to be the case also when short sequences, e.g. two-letter sequences, are abbreviated. Thus, a large part of the abbreviations in the examined manuscripts are used in lexically conditioned abbreviations, rather than graphemic-phonemic abbreviations, even when short sequences are abbreviated. Manuscripts containing a large amount of skaldic poetry are less abbreviated than those containing mainly prose. The poetic vocabulary in this type of poetry is archaic and generally marked by infrequency, and such words appear to have been abbreviated less. Possibly the scribes looked upon abbreviations as disturbing readability, and thus not to be used excessively in uncommon and obscure words.

Frequency is not the only factor triggering the use of abbreviations. For instance, it seems that very frequent endings, e.g. *-ir* and *-ar*, are abbreviated irrespective of word type. This is a case of the graphemic-phonemic abbreviation type, and it can lead to infrequent words being abbreviated as well. Furthermore, the structure of the words must in some way be a determining factor in the use of abbreviations. Obviously, the words must contain a sequence that can be abbreviated in a relatively unambiguous way. This leads to some highly frequent words not being abbreviated (e.g. *ráð*). Longer words that have an established abbreviated form are almost never written out in full, for instance the different forms of the lemmas *segja* and *konungr*.

The hypothesis that Icelandic manuscripts are more heavily abbreviated than the Norwegian ones is also confirmed and quantified in our investigation. Furthermore, as stated above, a clear chronological development towards an increased abbreviation tendency cannot be discerned in the Norwegian material. The earliest manuscript, AM 619 4to,

comes very close to the latest one, AM 78 4to. A comparison with one Old Swedish manuscript, Cod. Holm. B 59, from approximately the same time as the Norwegian material, showed that the abbreviation tendency was still considerably stronger in the Norwegian manuscripts than in the Old Swedish one.

Our investigation goes as far as the available empirical material allows. An important task for future research on the habits of the medieval scribes is to make available more manuscripts in digital form, preferably in a format where both abbreviation and expansion are available. With more data we will be able to elaborate and refine the results of this investigation.

Bibliography

- Bandlien, Bjørn. 2014. "Sir Snara Asláksson owns me": The historical context of Uppsala De la Gardie 4–7. In: Karl G. Johansson and Else Mundal (eds.), *Riddarasögur. The Translation of European Court Culture in Medieval Scandinavia*, pp. 245–271. Oslo: Novus.
- Driscoll, Matthew. 2009. Marking-Up Abbreviations in Old Norse-Icelandic Manuscripts. In: Maria Grazia Saibene and Marina Buzzoni (eds.), *Medieval Texts – Contemporary Media: The Art and Science of Editing in the Digital Age*, pp. 13–34. Como: Ibis.
- Guðvarður Már Gunnlaugsson. 2007. *Sýnisbók íslenskrar skriftar*. 2. ed. Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum.
- Haugen, Odd Einar. 2002. The Development of Latin Script I: In Norway. In: Oskar Bandle et al. (eds.), *The Nordic Languages. An International Handbook of the History of the North Germanic Languages I*, pp. 824–832. (Handbücher zur Sprach- und Kommunikationswissenschaft 22.1.) Berlin and New York: Walter de Gruyter.
- Haugen, Odd Einar. 2018. Høgmellomalderen (1050–1350). In: Helge Sandøy and Agnete Nesse (eds.), *Norsk språkhistorie IV*, Tidslinjer, pp. 198–292. Oslo: Novus.
- Hreinn Benediktsson. 1965. *Early Icelandic script as Illustrated in Vernacular Texts from the Twelfth and Thirteenth Centuries*. (Íslenzk handrit. Series in folio 2.) Reykjavík.

- Johansson, Karl G. 1997. *Studier i Codex Wormianus. Skrifttradition och avskriftsverksamhet vid ett isländskt skriptorium under 1300-talet*. (Nordistica Gothoburgensia 20.) Göteborg: Göteborgs universitet.
- Kjeldsen, Alex Speed. 2013. *Filologiske studier i kongesagahåndskriftet Morkinskinna*. (Bibliotheca Arnamagnæana. Supplementum 8.) Copenhagen: Museum Tusculanum.
- de Leeuw van Weenen, Andrea. 1993. *The Icelandic Homily Book. Perg. 15 4^o in the Royal Library, Stockholm*. (Íslensk handrit. Series in quarto 3.) Reykjavík: Stofnun Árna Magnússonar á Íslandi.
- Lindblad, Gustaf. 1954. *Studier i Codex regius av Äldre Eddan*. I–III. (Lundastudier i nordisk språkvetenskap A:10.) Lund: Gleerup.
- The Menota handbook: Guidelines for the electronic encoding of Medieval Nordic primary sources*. Gen. ed. Odd Einar Haugen. Version 3.0. Bergen: Medieval Nordic Text Archive, 2019. <<http://www.menota.org/handbook.xml>> (last accessed 18-02-2025).
- Mørck, Endre. 2018. Seinmellomalderen (1350–1536). In: Helge Sandøy and Agnete Nesse (eds.), *Norsk språkhistorie IV*, Tidslinjer, pp. 293–356. Oslo: Novus.
- Mårtensson, Lasse. 2011. *Studier i AM 557 4to. Kodikologisk, grafonomisk och ortografisk undersökning av en isländsk sammelhandskrift från 1400-talet*. (Rit 80.) Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum.
- ONP = *Ordbog over det norrøne prosasprog. Dictionary of Old Norse Prose*. University of Copenhagen, Department of Nordic Studies and Linguistics. <<http://onp.ku.dk/onp/onp.php>> (last accessed 18-02-2025).
- Stefán Karlsson. 2002. The Development of Latin Script II: In Iceland. In: Oskar Bandle et al. (eds.), *The Nordic Languages. An International Handbook of the History of the North Germanic Languages I*, pp. 832–840. (Handbücher zur Sprach- und Kommunikationswissenschaft 22.1.) Berlin and New York: Walter de Gruyter.
- Wills, Tarrin. 2024. Abbreviation in Old Norse Manuscripts – a quantitative study. *Gripla* 35, pp. 69–104.

Webpages

The Emroon homepage: <<https://www.emroon.no>>

The Menota homepage: <<https://clarino.uib.no/menota>>

Sammandrag

Det har ofta framhållits att bruket av förkortningar är ett kännetecknande drag för de fornvästnordiska handskrifterna, särskilt de isländska. Denna artikel behandlar bruket av förkortningar i företrädesvis fornisländska handskrifter, under jämförelse med forn norska handskrifter och även en fornsvensk handskrift. Materialet utgörs av digitala utgåvor av handskrifter eller delar av handskrifter, till största delen tagna från de digitala arkiven hos Menota och Emroon. De isländska handskrifterna i undersökningen sträcker sig från ca 1280 till inledningen av 1500-talet, medan de norska sträcker sig mellan 1200–1350, men med en koncentration till 1270–1325. Den fornsvenska handskriften är från ca 1280. Bland de isländska handskrifterna kan man se en gradvis utveckling över tid mot kraftigare förkortning; i handskriften från 1280 är 33 % av orden förkortade, medan handskriften från början av 1500-talet har 62 % av orden förkortade. En sådan utveckling kan inte iakttagas bland de norska handskrifterna, vilka dock har tillkommit under en mer begränsad tidsperiod. En annan tendens är att graden av förkortning är lägre i poesi, med en stor andel ovanliga ord jämfört med vanlig prosa. Man kan också se en påtaglig regional variation i bruket av förkortningar. De isländska handskrifterna är betydligt mer förkortade än de norska; de förstnämnda varierar i spannet 33–62 % förkortade ord, medan de sistnämnda varierar i spannet 10–26 % förkortade ord. Samtidigt är också de norska handskrifterna mer förkortade än den undersökta fornsvenska handskriften. Den har 5,6 % förkortade ord, en avsevärt lägre andel än de norska handskrifterna.

Lasse Mårtensson
 Uppsala University
 Department of Scandinavian Languages
 Box 527
 SE-751 20 UPPSALA
 lasse.martensson@nordiska.uu.se
 ORCID iD 0000-0001-5072-4961

LASSE MÅRTENSSON, EVA PETTERSSON AND VETURLIÐI ÓSKARSSON

Eva Pettersson
Uppsala University
Department of Linguistics and Philology
Box 635
SE-751 26 UPPSALA
eva.pettersson@lingfil.uu.se
ORCID iD 0000-0002-1447-4501

Veturliði Óskarsson
Uppsala University
Department of Scandinavian Languages
Box 527
SE-751 20 UPPSALA
veturlidi.oskarsson@nordiska.uu.se
ORCID iD 0000-0003-0434-0611