TARRIN WILLS

ABBREVIATION IN OLD NORSE MANUSCRIPTS

– a quantitative study

Introduction

Abbreviation in Icelandic and Norwegian Manuscripts

Old Norse manuscripts from Iceland and Norway employ extensive abbreviation[1] in representing text. The accepted picture of abbreviation in early Norse manuscripts is expressed by Hreinn Benediktsson in *Early Icelandic Script* and can be paraphrased as follows: a complex system of abbreviation of Latin emerged on the Continent in the eleventh and twelfth centuries, around the time that Icelandic script was developing. This system was not used much in other vernacular writing systems, but it was applied extensively in writing Norwegian and Icelandic, adapted and integrated into those scripts, and then developed further. Both Icelandic and Norwegian use a great deal of abbreviation, but in Icelandic script it is particularly distinctive, and increases in the late twelfth and thirteenth centuries.[2] This picture had already been established by Kålund[3] and discussed in further detail by Seip.[4]

Hreinn notes that this development runs in parallel with the use of non-phonemic ligatures and is driven by 'graphic expediency and econo-

1 This paper uses 'abbreviation' in both the abstract (as here) and concrete (often in the plural) sense of an abbreviation mark, the former being the original usage in English according to the OED.

2 Hreinn Benediktsson, *Early Icelandic Script as Illustrated in Vernacular Texts from the Twelfth and Thirteenth Centuries*, Íslenzk Handrit/Icelandic Manuscript Series in Folio II (Reykjavík: The Manuscript Institute of Iceland, 1965), 86–7.

3 Kristian Kålund, *Palæografisk atlas: Ny serie. Oldnorsk-islandske skriftprøver c. 1300-1700*. (København and Kristiania: Gyldendal, 1907), vii–viii.

4 Didrik Arup Seip, *Palæografi. B: Norge og Island*, Nordisk Kultur 28:B (Stockholm: Albert Bonnier, 1954), 30–1, 59.

my', pointing to the First Grammatical Treatise. That treatise is systematic in its phonemic analysis but applies principles of economy in various ways, such as the representation of geminates as a single letter and a particular discussion of abbreviation marks.

Hreinn makes a number of observations that can be presented as hypotheses:

1. Icelandic and Norwegian medieval manuscript writing uses abbreviation more extensively than other vernaculars.
2. Icelandic manuscript writing uses more abbreviation than Norwegian.
3. The amount of abbreviation increases in the first centuries of manuscript writing in Iceland.
4. Abbreviation is driven by economy, that is, saving physical space on the manuscript page.

Points 1–3 in particular are broadly consistent with the earlier scholarship of Kålund and Seip as well as later studies by Haugen[5] and Stefán Karlsson, for example.[6] The first three points are observations of measurable phenomena but are not supported by explicit quantitative analysis. It is the aim of the present study to test and refine these observations on digital corpora and by doing so give some insight into the fourth hypothesis.

Testing the first of these hypotheses requires a body of non-Norse digital transcriptions which can be compared with Norse ones, that is, digital texts which mark up abbreviations in a similar way to the standards in Norse digital editing. As the Text Encoding Initiative (TEI) has become the *de facto* standard for digital editing in the humanities in recent decades, this should be possible, and some studies have been made of Latin and vernacular corpora to compare abbreviation usage. Honkapohja provides a

5 Odd Einar Haugen, "The Development of Latin Script I: in Norway," in *The Nordic Languages. An International Handbook of the History of the North Germanic Languages* 1, ed. Oskar Bandle et al. (Berlin & New York: Walter de Gruyter, 2002), 830.

6 Stefán Karlsson, "The Development of Latin Script 2: in Iceland," in *The Nordic Languages. An International Handbook of the History of the North Germanic Languages* 1, ed. Oskar Bandle et al. (Berlin & New York: Walter de Gruyter, 2002), 835.

very useful review of these studies,[7] and I draw upon it heavily in the following part of the literature review.

Existing studies measure the phenomenon by the ratio of the number of words that are abbreviated to the total word count. For Latin, this is sometimes more than 50 per cent of words, for English up to 30 per cent[8] and a similar figure for French.[9] The amount of abbreviation in these languages decreased from the fourteenth to the sixteenth century, and is almost absent in the earliest printed works.[10] The aforementioned studies focus on specific genres that vary greatly in prestige and consequently the resources used for writing. The status of scribes and patrons, as well as the manuscripts themselves (e.g. as measured by size and margins) seems to have had an impact on abbreviation rates in Latin and vernacular manuscripts.[11] Higher-status manuscripts tend to use less abbreviation, and more utilitarian ones much more. Other studies have noted the inverse relationship between manuscript size and the amount of abbreviation:[12] smaller manuscripts abbreviate more, perhaps because they are generally economising on the use of the page surface.

For Old Norse, the amount of abbreviation has been only a matter of speculation in the published literature, expressed for example as a maximum of one third of words abbreviated,[13] or more than medieval Latin,[14]

7 Alpo Honkapohja, "Digital Approaches to Manuscript Abbreviations: Where Are We at the Beginning of the 2020s?," *Digital Medievalist* 14 (2021) DOI: http://doi.org/10.16995/dm.88.

8 Alpo Honkapohja and Aino Liira, "Abbreviations and Standardisation in the Polychronicon: Latin to English, and Manuscript to Print," in *The Multilingual Origins of Standard English (MOSTE)*, ed. Laura Wright (Berlin: De Gruyter, 2020), 269–316. DOI: http://doi.org/10.1515/9783110687545-010, p. 281.

9 Emilie Cottereau-Gabillet, "Revealing Some Structures and Rules of Book Production (France, Fourteenth and Fifteenth Centuries)," in *Ruling the Script in the Middle Ages. Formal Aspects of Written Communication (Books, Charters, and Inscriptions)*, ed. Sébastien Barret, Dominique Stutzmann, and Georg Vogeler, Utrecht Studies in Medieval Literacy 35 (Turnhout: Brepols, 2016), 129–63.

10 Honkapohja and Liira, "Abbreviations and Standardisation," 279, 282–3.

11 E.g. Cottereau-Gabillet, "Revealing Some Structures."

12 Alpo Honkapohja, "'Latin in Recipes?' A Corpus Approach to Scribal Abbreviations in 15th-Century Medical Manuscripts," in *Multilingual Practices in Language History: English and beyond*, ed. Päivi Pahta, Janne Skaffari and Laura Wright (Berlin: De Gruyter, 2018), 249.

13 Haraldur Bernharðsson and Odd Einar Haugen, "Chapter 6. Abbreviations" in *Menota Handbook*, ed. Odd Einar Haugen, 3rd ed. (Bergen: Medieval Nordic Text Archive, 2019).

14 Matthew James Driscoll, "Marking up Abbreviations in Old Norse-Icelandic

by inference, more than half of words abbreviated. The observations about the relationship between abbreviation in Icelandic, Norwegian and Continental vernacular manuscripts imply that Iceland's cultural relationship with Europe and Norway diverged in the course of the Middle Ages. That is, a part of the important cultural practice of book production was already distinctive when it started in Iceland (at least in relation to other vernaculars), and rather than converging over time with European vernacular practices as Iceland came more into contact with them, it in fact became even more distinct. Even without a direct comparison, one can assume that an increase in abbreviation within the corpus suggests a divergence from European tradition, in which abbreviation decreased over the same period.

The second and third hypotheses will be tested in this study to establish a quantitative foundation for them. While Hreinn's study is, by definition, restricted to early Icelandic script, the body of data now available allows us to extend the diachronic observation of the twelfth and thirteenth centuries to the entire period of Icelandic manuscript production in order to understand how abbreviation practice changed over a longer period of time. This then provides a possible observation about the potential divergence from and convergence with European vernacular practice over time.

The fourth hypothesis, that abbreviation is driven by economy, implies motivation, which is a very difficult thing to establish when only the results of a human activity are available. However, parchment was expensive, relying on slaughtering livestock which may have had other uses, for example, wool and milk production.[15] Parchment and manuscript production were also labour-intensive, and both diverted resources from economic and subsistence needs (leather production and farm work, for example). It can be inferred that there would be a motivation to economise on both materials and labour in manuscript production by reducing the amount of material and time used to write texts.[16] Conversely, manuscripts, precisely because of their expense, were likely also status symbols when expansive and richly decorated. In either case, the amount of parchment used is re-

Manuscripts," in *Medieval Texts – Contemporary Media: The Art and Science of Editing in the Digital Age*, ed. Maria Grazia Saibene and Marina Buzzoni (Pavia: Ibis, 2009), 13–34.

15 It should be noted that excess male calves are a normal by-product of dairy farming, and vellum in particular can be understood thus as a by-product of milk production, albeit requiring additional labour and materials. Calf skins, however, presumably had potential uses other than as vellum for manuscripts.

16 Anthony G. Petti, *English Literary Hands from Chaucer to Dryden* (London: Arnold, 1977), 22.

lated to wealth and economy. I take economy of parchment usage therefore as an assumption in this study, which allows us to use it as a measure of abbreviation: how much physical space was saved by abbreviating text gives us an important measure of the extent of abbreviation.

The present study focuses in particular on the use of abbreviation in writing poetry and prose in manuscripts. In order to undertake this study in context, it needs to be established whether there was a difference in the practice of abbreviation between poetry and prose. It can be inferred that poetry was more difficult to understand than prose and required a slightly different process in copying.[17] There has been no published study comparing the use of abbreviation in poetry and prose in Icelandic manuscripts, although Kjeldsen in personal communications has noted that he has observed a marked difference between poetry and prose in the amount of abbreviation in *Morkinskinna* (GKS 1009 fol.). The middle part of this study therefore attempts to compare the extent of poetic and prose abbreviation in manuscripts where both occur, before proceeding to the analysis of the abbreviation of skaldic poetry over time.

Representation of Abbreviation and Expansion

The practices involved in editing manuscript texts have been driven over the centuries by sometimes conflicting needs and constraints: technological constraints in particular limit the ability to represent the uniqueness of each handwritten document, and other needs have put differing emphasis on standardising a text to make it comparable to other texts or accessible to readers who are familiar with the language in its reconstructed form but not the manuscript orthography. Normalisation is important to making early texts accessible and is often essential to linguistic, stemmatological, stylometric and other types of analysis.

Normalisation and expansion of abbreviation removes the possibility of digitally examining abbreviation,[18] albeit only when expansion is silent. While expansion and further normalisation is used for various practical and research purposes, many of the works that are critical of this practice assume that abbreviation is removed because it is considered accidental

17 E.g. Alex Speed Kjeldsen, *Filologiske studier i kongesagahåndskriftet Morkinskinna*, Bibliotheca Arnamagnaeana. Supplementum 8 (Copenhagen: Museum Tusculanum, 2011), 777–8, 883.

18 Honkapohja, "Digital Approaches to Manuscript Abbreviations."

(although all language is in a sense accidental) or trivial. However, where abbreviation is marked in the form of expansion, it provides both a means of identifying abbreviated words and, to a certain extent, an understanding of the amount of apparent text that has been abbreviated. Further, in his review Honkapohja does not discuss the various techniques that preserve abbreviated, expanded and linguistically normalised versions of the same text. A further review of editing technologies is therefore required to determine whether different types of editions, rather than only those that record abbreviated forms, can be used to analyse abbreviation use.

Printed and Simple Diplomatic Editions

Early printed editions of Old Norse expanded abbreviations silently and often normalised the text to a certain extent.[19] This was partly no doubt due to typographic constraints, together with a focus on making the texts accessible rather than on fidelity to the manuscripts. Towards the end of the nineteenth century, printed diplomatic editions began to present the text with expansions of abbreviations marked using italics.[20] This was probably driven to a certain extent by the series published by Samfund til udgivelse af gammel nordisk literatur (STUAGNL), the Norse counterpart to the Early English Text Society, itself a body founded to prepare editions for the future Oxford English Dictionary. STUAGNL began this practice in most of the editions in its first year of publishing (1880), and it quickly became standard. This practice in Old Norse diplomatic editing is now ubiquitous: abbreviations are almost always expanded using italics unless the text is normalised.

One of STUAGNL's early editions (Dahlerup's 1880 edition of *Ágrip* – volume 3 in the series[21]) attempted, however, to reproduce the abbreviations in their unexpanded forms and even included a facsimile of one of

19 Cf. Gottskálk Jensson, "Udgivelse af norrøn litteratur indtil 1772," in *Dansk Editionshistorie 2: Udgivelse af norrøn og gammeldansk litteratur*, ed. Britta Olrik Frederiksen (Copenhagen: Museum Tusculanum, 2021), 48.

20 E.g. Eugen Kölbing, ed., *Elis saga ok Rosamundu* (Heilbronn: Henninger, 1881); Carl af Petersens, ed., *Jómsvíkinga saga efter Arnamagnæanska handskriften N:o 291. 4:to*, STUAGNL 7 (Copenhagen: S. L. Møller, 1882). Compare Carl af Petersens, ed., *Jómsvíkinga saga (efter Cod. AM. 510, 4:to) samt Jómsvíkinga drápa*, (Lund: C. W. K. Gleerup, 1879), where expansion is silent.

21 Verner Dahlerup, *Ágrip af Noregs konunga sögum*, STUAGNL 3 (Copenhagen: S. L. Møller, 1880).

the manuscript's leaves. The typeset lines are widely spaced in order to accommodate the interlinear marks typical of abbreviation. This practice was employed sparingly in print, however, no doubt due to typographic challenges and the extra space required on the page.

Some of the digital corpora used below, including the Skaldic Project's transcription corpus and some Menota XML files, have used the technique of marking abbreviation expansions, rather than abbreviations themselves, extensively. The practice of representing expansions using italics gives some information about abbreviation: which words are abbreviated, and in addition, the extent to which words are abbreviated. Analysing this data relies on the ability to extract this information unambiguously in a digital form.

TEI XML

The first major release of the Text Encoding Initiative's (TEI) guidelines, TEI P3 (1994–1999),[22] included a means of digitally representing both abbreviations and expansions in a simple data structure (either abbreviation with an expansion attribute, or vice-versa), and these methods were adopted unchanged in the first XML version of TEI (P4, 2002). Wills, for example, used this method to produce digital and printed versions of Old Norse manuscripts which could be read in either their abbreviated or expanded form.[23]

The next and current version of TEI (P5, 2007-)[24] generally removed unstructured character data from attributes, resulting in a slightly more complicated encoding but more possibilities for adding additional information about abbreviations and expansions. The first decade of this century produced a number of different proposals for methods of encoding abbreviations and their corresponding expansions, focusing on particular problems of the sometimes complex relationship between abbreviation

22 C. M. Sperberg-McQueen and Lou Burnard, eds., *Guidelines for Electronic Text Encoding and Interchange* (Chicago and Oxford: TEI P3 Text Encoding Initiative, 1994).

23 Tarrin Wills, The Foundation of Grammar: An Edition of the First Section of Óláfr Þórðarson's Grammatical Treatise, PhD Thesis (University of Sydney, 2001).

24 TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, Version 4.7.0. (TEI Consortium, 16 November 2023): http://www.tei-c.org/Guidelines/P5/ (accessed 23 March 2024).

marks and what they represent. Honkapohja gives a detailed review of digital manuscript abbreviations using strict TEI P5 and recommends a markup system of this type:[25]

```
<w>
<choice>
<abbr>magn<am>&#42863;</am></abbr>
<expan>magn<ex>us</ex></expan>
</choice>
</w>
```

The main variation in this method is whether the whole word is treated in an abbreviated and expanded form, or the abbreviations only; and whether the abbreviation mark (here encoded as the Unicode LATIN SMALL LETTER CON) itself is marked up with the <am> tag, which assists in identifying marks used for abbreviations. This type of markup is the basis for various studies and facilitates the digital counting of abbreviated words, identifiable by the presence of the <abbr> element. The presence of any of the four element types used here (<abbr>, <expan>, <am>, <ex>) in a word would indicate an abbreviated word.

The method of treating the process of abbreviation at the word level (logographic) rather than the abbreviation marks themselves solves an issue where the abbreviation marks do not correspond closely to the putative expansion (e.g. '.e.' > 'eða', where the first dot is an abbreviation mark but does not have a clear relationship to the expanded text).

This type of encoding has formed the basis of many studies of abbreviations in Latin and vernacular manuscripts which we will use for comparison. No possibility is presented in the papers referenced in this article for adding a normalised version in pure TEI, making it difficult to compare abbreviation of particular words across manuscript versions.

Menotic TEI

The *de facto* standard which has emerged in Old Norse textual editing is that described in the Menota Handbook.[26] Menota uses a modified

25 Honkapohja, "Digital Approaches to Manuscript Abbreviations."

26 Odd Einar Haugen (ed.), The Menota Handbook: Guidelines for the Electronic Encoding

version of TEI P5 with a separate namespace ('me:') in order to encode a slightly different structure for abbreviations, expansions and normalisations. Menota has become the primary method used in the digital editing of Old Norse manuscripts, and its archive at the time of writing contains some ninety documents with over two million words.

Menota's method is to separate the abbreviated and expanded forms of words (tokens) into two 'levels': 'facsimile' and 'diplomatic' respectively. The facsimile level represents the characters, including abbreviations, as they appear on the page, and the diplomatic level corresponds in terms of abbreviation to the traditional print diplomatic editions. These are semantically similar to the markup advocated by Honkapohja and others, but allow for other non-linguistic features to be separated into transcription 'levels'. Menotic abbreviation markup can be converted without information loss to standard TEI where abbreviations are encoded at the word level.

The example abbreviated word above can be represented as follows:

```
<w>
<choice>
<me:facs>magn<am>&#42863;</am></me:facs>
<me:dipl>magn<ex>us</ex></me:dipl>
<me:norm>Magnús</me:norm>
</choice>
</w>
```

The manuscript variation itself is encoded, along with a putative expansion of the abbreviations, in addition to a normalised rendering of the language of the manuscript, which can be compared with other texts and versions of it. Although Honkapohja mentions Menota, the project's particular method of marking up abbreviation and expansion is not mentioned in the review of encoding techniques. This gives the mistaken impression, when taken in conjunction with the discussion of the problems of normalisation,[27] that normalisation must be abandoned in order to allow for the digital investigation of abbreviation. The Menota model in fact

of Medieval Nordic Primary Sources, Version 3.0 (Bergen: Medieval Nordic Text Archive, 2019).

27 Honkapohja, "Digital Approaches to Manuscript Abbreviations," §§4-5.

avoids the trade-offs of normalisation as described by Honkapohja. It does, however, produce additional code, as all words are encoded with multiple transcription levels or choices, not only those that contain abbreviations.

In Menota editions abbreviations can be identified by the <am> element if it is used to mark abbreviations at the 'facsimile' level, and they can always be identified by the presence of the <ex> element at the 'diplomatic' level. In practice, no Icelandic manuscripts and only one Norwegian manuscript in the Menota archive lack the diplomatic level. The <ex> element can therefore be used to identify abbreviations in almost all cases.

In an ideal situation, no information is discarded, but all three types of information are recorded: the letters, abbreviations, spacing and other features of the physical manuscript page; the putative expansions based on the editor's understanding of the scribe's normal orthography and use of abbreviation marks; and the normalisation, which represents the editor's understanding of the well-established reconstructed language of the time and place of the manuscript, and which allows comparisons with other manuscripts that use the same language but differing orthography. In practice, however, recording and in particular checking such detailed structures is very time-consuming, although newer tools such as MenotaBlitz and MenotaG (menotag.ku.dk) promise to make this process easier.

Until we have a large body of comparable material marking both abbreviations and expansions, the simpler approach of marking expansions (the 'diplomatic' level) provides a potentially larger and more diverse corpus for investigating abbreviation quantitatively. I therefore make use of expanded diplomatic texts where the expansion is marked up. This requires a method to measure abbreviation and an examination of the assumptions that lie behind that method.

Types of Abbreviation

Examining the types of abbreviation in the available corpora gives an overview of how abbreviations are expanded and therefore the relationship between the script and the text. The typology of abbreviation in Old Norse manuscripts was established by Kålund,[28] and is used with some

28 Kålund, *Palæografisk Atlas*, viii–x.

variation by Seip[29] and Hreinn Benediktsson.[30] It is summarised more recently in the Menota *Handbook*. Although there are small differences in how different abbreviation practices are categorised by earlier scholarship, they follow largely the categories outlined in the Menota *Handbook*, as summarised here:[31]

1. Suspension: the word is abbreviated by removing letters from the end and replacing them with a punctuation-like mark.
2. Contraction: the word is abbreviated by removing letters from the middle of the word and often indicated by an interlinear mark such as a horizontal bar.
3. Interlinear marking: the word is abbreviated by removing letters from the baseline and replacing it with an interlinear abbreviation mark, usually a letter implying a combination of that letter with r, v or a.
4. Baseline brevigraphs: Special marks on the baseline that do not consist of ordinary letters but represent letter combinations, in particular the Tironian notae.

Examples of each are shown in Table 1.

Abbreviated word	**Type of abbreviation**	**Expansion of word**
ſ.	Suspension	s*onr*
ſſ.	Suspension	s*ynir*
k̄ḡr	Contraction	k*onun*gr
lð	Contraction	l*an*d
e̊	Interlinear mark	e*r*
þ̄	Interlinear mark	þ*at*
þ̄	Interlinear mark	þ*ar*
7̷	Brevigraph	*ok*
Ψ	Brevigraph	*maðr*

Table 1: Abbreviation examples.

29 Seip, *Palæografi B*, 61–2.
30 Hreinn Benediktsson, *Early Icelandic Script*, 85.
31 Haraldur Bernharðsson and Haugen, "Abbreviations."

In the context of our present study, all of these methods involve fewer strokes of the pen on the page than writing out the corresponding unabbreviated text, and all involve a reduction in the number of baseline letters written on the manuscript page. The horizontal space used to write the text can therefore accommodate more text, while the amount of vertical space remains the same.

The marked-up letters (italics corresponding to expansion tagging) are inferred from the abbreviation and sometimes context. The abbreviation may include a letter indicated in the expansion, such as the superscript <r> in <þ>, but the general principle is that if a letter is written interlineally or not written at all, it is marked up in the expansion, for example with italics in print. In all cases except the brevigraphs, the non-marked-up (visually or in code) letters correspond to the letters or letter-like characters that occupy the baseline of the text. The amount of economy of the abbreviation can therefore be measured in most cases as the relationship between the number of letters that have not been marked up as expansions and the total number of letters including the marked-up expansions.

The exceptions here are the 'ff.'-type and the brevigraphs. In the 'ff.' case only one of the letters written on the baseline is included in the part of the text not marked as an expansion. These instances are relatively rare, however, and the difference in the resulting ratio between abbreviated and expanded width is in any case not great. Brevigraphs are also expanded with the full word marked as the expansion, even though the manuscript contains a baseline character. All but the Tironian notae (⁊), however, are relatively rare. The notae are uniformly expanded as *ok* (occasionally *og*) in the corpora used here, and therefore can be easily identified in the digital text as marked up <ex>ok</ex>, <ex>oc</ex> or <ex>og</ex>, with a high degree of confidence that this expansion corresponds to a single letter-like mark on the manuscript baseline.

Measures of Abbreviation

Proportion of Abbreviated Words

Where words are marked up (tokenised) and there is markup which identifies words with abbreviations, abbreviation can be measured by the proportion of words that are abbreviated in a manuscript. This measure is

the most commonly used in previous scholarship.[32] These studies normally examine corpora with TEI or similar markup. Where the text is marked up so that all words are identifiable and contain abbreviation markup for the whole word where abbreviated, one can simply count the number of these two tags to get a percentage of words abbreviated. This ignores instances where a word has two abbreviations such as e.g. 'mꝛ' for 'mæl*ir*'.

This measure has been applied very sparingly to Old Norse manuscript corpora but is implied by the Menota *Handbook* chapter 6: 'In some Icelandic manuscripts, as many as a third of the words may be abbreviated',[33] although this is not based on systematic measurement. A variation on this measure for Old Norse is Kjeldsen, who examines a shift within the *Morkinskinna* scribe A's use of abbreviations for common words.[34] Here the measurement is in the form of the number of abbreviations observed per hundred words in the text, in which case a word may be counted twice if it includes more than one abbreviation.

As Menota-style TEI files have both words and abbreviations marked up (either or both as abbreviation marks and expansions), this measure can be applied to the Menota corpus. It is, however, less easy to apply this measure to corpora that are not tokenised and cannot be reliably tokenised.

A Measure of Abbreviation as Economy of Text

In a corpus where expansions only are marked up and there is no tokenisation, a different measure of abbreviation is needed. I also aim here specifically to measure economy, that is the reduction in page surface usage realised by abbreviation. This measure should then reflect the amount of page surface saved by the scribe in abbreviating the text.

To illustrate how the marked-up expansions can be used to measure abbreviation economy, I use a line from AM 748 I b 4to as an example, chosen because of its many abbreviations (Figure 1).

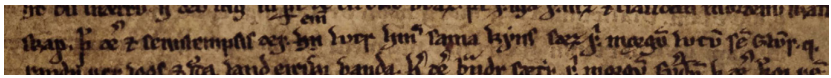

Figure 1: AM 748 I b 4to 12r/25.

32 E.g. Cottereau-Gabillet, "Revealing Some Structures"; Honkapohja, "Latin in Recipes?," etc.

33 Haraldur Bernharðsson and Odd Einar Haugen, "Chapter 6. Abbreviations" in *Menota Handbook*, ed. Odd Einar Haugen, 3rd ed. (Bergen: Medieval Nordic Text Archive, 2019).

34 Kjeldsen, *Morkinskinna*, 780–2.

The line of text in Figure 1 is transcribed by the present author:

> þ*at ær ok* semilempsis æf ein*n* lvtr hin*s* sama kyns sæz f*yrir* mǫrgv*m* lvtv*m* se*m* Glv*m*r q*vað*·

The marked-up expansions can be encoded thus (ignoring the corrected word):

> þ<ex>at</ex> æ<ex>r</ex> <ex>ok</ex> semilempsis æf ein<ex>n</ex> lvtr hin<ex>s</ex> sama kyns sæz f<ex>yrir</ex> mǫrgv<ex>m</ex> lvtv<ex>m</ex> se<ex>m</ex> Glv<ex>m</ex>r q<ex>vað</ex>·

The word 'hin*s*' does not contain an abbreviation in the strict sense: all letters are written in full, although the final one is written over the last letter. In this measure, however, which seeks to measure page surface usage, it is treated as abbreviation, as it abbreviates the horizontal length of the line.

As spaces do not affect abbreviation economy, these are removed, along with any other non-abbreviation-related tags. Expanded *ok* is converted so that it only has one letter expanded (bold), reflecting the fact that it occupies in its abbreviated form one character on the baseline:

> þ<ex>at</ex>æ<ex>r</ex>**o<ex>k</ex>**semilempsisæfein<ex>n</ex>lvtrhin<ex>s</ex>samakynssæzf<ex>yrir</ex>mǫrgv<ex>m</ex>lvtv<ex>m</ex>se<ex>m</ex>Glv<ex>m</ex>rq<ex>vað</ex>·

This string of text is used to calculate the relative economy of abbreviation. The first value in the calculation is the character length of the string with the expansions removed, that is, characters corresponding to letters appearing on the baseline of the manuscript line, i.e. 55 characters:

> þæosemilempsisæfeinlvtrhinsamakynssæzfmǫrgvlvtvseGlvrq·

This is compared with the length of the string with only the expansion tagging removed, i.e. the reconstructed expanded text totalling 72 characters:

þatæroksemilempsisæfeinnlvtrhinssamakynssæzfyrirmǫrgvml-
vtvmsemGlvmrqvað·

The measure used is the number of letters in the expansions relative to the total number of letters including expansions. Here there are 17 letters in the expansions (72–55), i.e. 17/72 = 0.236, that is, the scribe on this line has economised by 23.6% baseline letters from the putative expanded form. For comparison, 11 of the 17 words are abbreviated (65%).

This measure requires a reconstruction on the part of the editor which, it could be argued, cannot be determined as confidently as the presence or absence of abbreviation marks, as is used in previous studies. Here, for example, I have expanded f with superscript i as f*yrir*, but f*yr* might also be possible. Other examples include þ*ei*ra/þ*ei*rra or k*onun*gr/k*ong*r. In practice, however, there are few such ambiguities, and I assume here that any differences largely cancel each other out or do not affect the overall results.

This technique has the advantage that it opens up the possibility of examining corpora that only have expansions marked up. This includes a large number of manuscript editions in the Menota archive that have a diplomatic level but no facsimile level, and the large corpus from the Skaldic Project, where expansions are marked up. Future studies could potentially draw on printed editions with italic expansions. In addition, it measures better abbreviation according to the fourth hypothesis deriving from Hreinn Benediktsson in the introduction, as it counts more directly the amount of horizontal space, measured in characters, saved by the process of abbreviation.

Where both measures (abbreviated word percentage and abbreviation economy) can be applied, the measures can be compared to determine the relationship between the two.

Comparing the Two Measures

The Menota archive contains a large body of manuscript transcriptions to which both measures of abbreviation can be applied. In Figure 2, these manuscripts are analysed to identify the percentage of abbreviated word tokens (horizontal axis), which is plotted against the abbreviation economy percentage value (vertical axis) for each manuscript (dot).

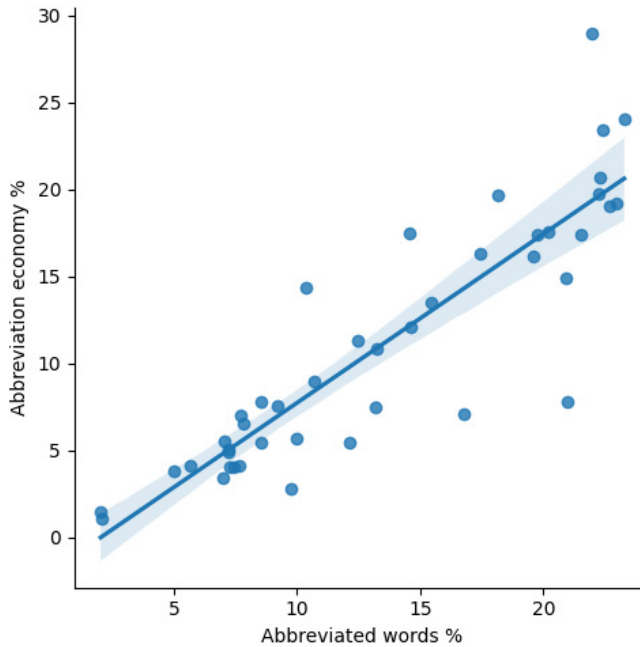

Figure 2: Menota manuscripts by abbreviation percentage and economy.

The small confidence interval here around the trendline indicates that the relationship is almost linear. This can also be expressed using the Pearson's correlation coefficient (ρ[35]). The correlation is expressed as a value between 1 (complete positive correlation) and −1 (complete negative correlation), with a value of 0 representing no correlation. Here Pearson's ρ is 0.89, indicating a very strong linear correlation between the two measures: that is, the two are closely correlated. The outliers in the diagram (dots further from the line of best fit) are small fragments. Removing the fragments with fewer than 3000 words gives us an even closer correlation (r = 0.95) between the two abbreviation measures. The average relationship between the two measures (m) is 1.26, or 1.18 for Icelandic manuscripts.

35 Karl Pearson, "Notes on Regression and Inheritance in the Case of Two Parents," *Proceedings of the Royal Society of London* 58 (1895), 240–2.

That is, for an Icelandic manuscript which has 20% of words abbreviated, the expected abbreviation economy would be 17% (20/1.18).

This measure of abbreviation economy is therefore closely related to the established measure of percentage of abbreviated words, allowing us to accurately compare corpora where only expansions are used with studies based on the percentage of abbreviated words. The present study will therefore use the abbreviation percentage measure where available to directly compare corpora that have used the established measure (proportion of abbreviated words), but use the abbreviation economy measure for corpora where only expansions are provided.

Abbreviation Marks by Frequency

The main method outlined above makes observations about the frequency of abbreviation types in order to justify its assumptions. As the main corpus (the Skaldic Project's transcription database[36]) to be used as the data for this study does not mark up abbreviations, only expansions, it is necessary to test whether this approach will produce a reasonably accurate overview of the amount of abbreviation in manuscripts. This in turn requires surveying a corpus where abbreviations are marked up to determine whether the inference here is valid.

Table 2 shows the twenty most common abbreviation marks found in <am> elements in Icelandic and Norwegian manuscripts in the Menota archive, where the element is used. Numerical entity references and unicode characters are resolved as Menota/MUFI entity names for consistency.[37]

36 Skaldic Poetry of the Scandinavian Middle Ages: https://skaldic.org.

37 Medieval Unicode Font Initiative: https://mufi.info.

Entity reference for abbreviation mark	Unicode name (or MUFI in italics)	Number of marks found in <am>
&bar;	COMBINING OVERLINE	51250
&er;	COMBINING ZIGZAG ABOVE	31759
&etslash;	*LATIN ABBREVIATION SIGN SMALL ET WITH STROKE*	10017
&apomod;	MODIFIER LETTER APOSTROPHE	7727
&rsup;	COMBINING LATIN SMALL LETTER R	6137
&ovlmed;	*COMBINING MEDIUM-HIGH OVERLINE WITH FIXED HEIGHT (FULL-WIDTH)*	4963
&isup;	COMBINING LATIN SMALL LETTER I	4658
&et;	TIRONIAN SIGN ET	3170
.	FULL STOP	3157
&inodotsup;	COMBINING LATIN SMALL LETTER DOTLESS I	2949
&combmacr;	COMBINING MACRON	2612
&combcomma;	COMBINING COMMA ABOVE RIGHT	1902
;	SEMICOLON	1830
·	MIDDLE DOT	1514
&ra;	COMBINING LATIN SMALL LETTER FLATTENED OPEN A ABOVE	1344
&asup;	COMBINING LATIN SMALL LETTER A	1108
&sem;	*LATIN ABBREVIATION SIGN SEMICOLON*	922
&osup;	COMBINING LATIN SMALL LETTER O	915
&rabar;	*COMBINING ABBREVIATION MARK SUPER-SCRIPT RA OPEN A FORM WITH BAR ABOVE*	671
&combdot;	COMBINING DOT ABOVE	354

Table 2: The twenty most common abbreviation marks in Menota Norwegian and Icelandic manuscripts.

These twenty abbreviation marks account for the overwhelming majority of abbreviations found in the corpus (138959 instances/142822 abbreviation marks in total = 97.3%).

There are a number of abbreviation marks that are essentially allographs or script variants. These include &er; and &combcomma; for the

tittle, and &et; and &etslash; for the Tironian nota, but this kind of variation is not relevant to the present study. The majority of the abbreviations are written above another character which is not included in the expansion or are punctuation marks occupying minimal horizontal space. The 'ff.' (for s*ynir*)-type abbreviation is not indicated by this method, but I assume in any case that this type of abbreviation is unusual enough not to significantly alter the results below.

Excursus: Space Usage of Punctuation and Letters

The assumption of the abbreviation economy method is that punctuation marks in abbreviations do not make a significant difference to the amount of horizontal space used by the scribe, as these are silently removed when the text is expanded. Punctuation is normally small, and physically measuring a very large number of punctuation marks and their spacing relative to letters would be laborious. However, we have at our disposal another dataset which can be used to measure these phenomena. The MenotaG framework[38] is a Menota-based model for editing and processing texts from manuscript images. It incorporates handwritten text recognition (HTR) tools for segmenting the images into lines. Words and punctuation tokens can be marked by editors on the manuscript images by vertically dividing the line outlines. These are stored as polygons using OpenGIS data structures and can be analysed with spatial tools.

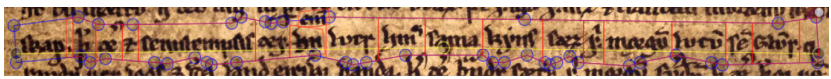

Figure 3: AM 748 I b 4to 12r/25 with MenotaG-generated outlines.

The HTR-generated line outline (blue dots) and baseline (yellow line) is shown along with the user-inputted token divisions (red outlines), which are also transcribed in both their abbreviated and expanded forms. HTR-generated token outlines tend to be an inaccurate reflection of the token width, and therefore user-inputted token divisions are used here. At this stage the system is being tested with three manuscripts of the Old

38 Cf. description in Tarrin Wills, "Asynchronous Linked Editing of Texts in Physical Objects," Digital Humanities in the Nordic and Baltic Countries, Reykjavík 27–31 May 2024 (DHNB 2024, forthcoming).

Icelandic *Third Grammatical Treatise*, with over 16,000 tokens marked in the three main manuscripts (AM 242 fol., AM 748 I b 4to and AM 757 a 4to), along with a few other manuscripts very partially segmented for testing purposes but which provide a point of comparison.

In order to physically measure the relative size of punctuation characters and word characters (including punctuation), a SQL query performs a number of look-ups and transformations. The query retrieves data from all images where word and punctuation tokens have been marked on the image as above. Combining characters (as defined by the MUFI project) and tags are removed from the token text, and spacing after the token is added where appropriate. The width of the rectangular bounds of the token polygon (in pixels) is compared with the number of characters of text (horizontal pixels per character). (HTR-generated token outlines tend to be an inaccurate reflection of the token width.) The ratio of pixels per punctuation token character to pixels per word token character is calculated. The average (weighted by the number of tokens on each image) of these from all images for a manuscript is then aggregated. The relative size is then calculated independently of the resolution of the images in pixels, which may vary even within manuscripts.

Ms siglum	**Relative punctuation width**	**Tokens**
AM 242 fol.	0.287	7456
AM 748 I b 4°	0.286	5881
AM 757 a 4°	0.328	3188
GKS 1009 fol.	0.396	270
AM 45 fol.	0.365	259
…		
Total	0.305	17826

Table 3: Relative width of letters and punctuation in MenotaG.

We see that, with a heavy focus on three manuscripts, punctuation characters on average occupy less than one third of the horizontal space of word characters. Using the data from the first three manuscripts in particular, we can conclude that the amount of horizontal space occupied by punctuation marks is therefore likely to be less than a third of that occupied

by letters. It is therefore reasonable to assume that punctuation does not contribute substantially to the use of the page surface. When a scribe uses a horizontally spacing punctuation mark in abbreviation, they are therefore adding the equivalent of one third of a character while saving on the unwritten, abbreviated characters.

Abbreviation in Icelandic and Norwegian Manuscripts in the Menota Archive

The measures established above are now applied to the available corpora of Old Norse manuscripts starting with abbreviation measured as the proportion of abbreviated words.

Using the Python programming language, I have written a script which scans the current Menota archive (as at 12 March 2024) and analyses the texts which are primarily Icelandic and Norwegian and have accessible XML files. Where a manuscript's text is found in multiple files, these are aggregated. XML files for a total of forty-four manuscripts have been examined. Menota's XML files are CC BY-SA-licensed and the editors for the files used are named as (in descending order of the number of tokens used in this study): Anna C. Horn, Karl Gunnar Johansson, Robert K. Paulsen, Fabian Schwabe, Nina Stensaker, Matteo Tarsi, The Bergen group (2), Beeke Stegmann, The Codex Regius project and Katarzyna Kapitan.

Using the XML data processed from Menota's archive, the extent of abbreviation was examined using the measures of number of abbreviations per word and proportion of words abbreviated. The total number of word tokens found in manuscripts that could yield results for the above process was 618,190. Of these, 138,893 were abbreviated in some way (22.5%), and the number of abbreviations in total was 143,336 (23.2 abbreviations per 100 words). The designation of either 'Norwegian' or 'Icelandic' is based on the designation in the archive catalogue and in some cases is misleading (e.g. Holm perg 4 fol., which has a mixture of apparently Icelandic and Norwegian hands[39]), but this affects very few data points in the following study.

39 Cf. e.g. Språksamlingane's introduction to the Menota edition at https://clarino.uib.no/menota/text/ menota/HolmPerg-4-fol (accessed 17 August 2024).

There was a very significant difference between the proportion of abbreviated words in Norwegian and Icelandic manuscripts, with Icelandic manuscripts having about 36% abbreviated words (56302/157415 words) compared with 18% for Norwegian (82591/460775 words). The proportion of abbreviated words in the Icelandic manuscripts in the Menota archive is therefore around double that of the Norwegian manuscripts. The latter in Menota are often legal documents, fragments and charters, which may distort these results to an extent. This nevertheless confirms Hreinn's observation about the difference in the amount of abbreviation in Icelandic and Norwegian manuscripts (hypothesis 2 above), which, if anything, is understated by him.

I will therefore treat Icelandic and Norwegian manuscripts separately, where possible, in the following analyses. Figure 4 plots the abbreviation percentages against the date (as the midpoint of a date range given in the Menota catalogue). There are very few manuscripts dated to after c. 1400 in the archive, making it difficult to examine diachronic changes after that point. These outliers in dating are therefore removed.

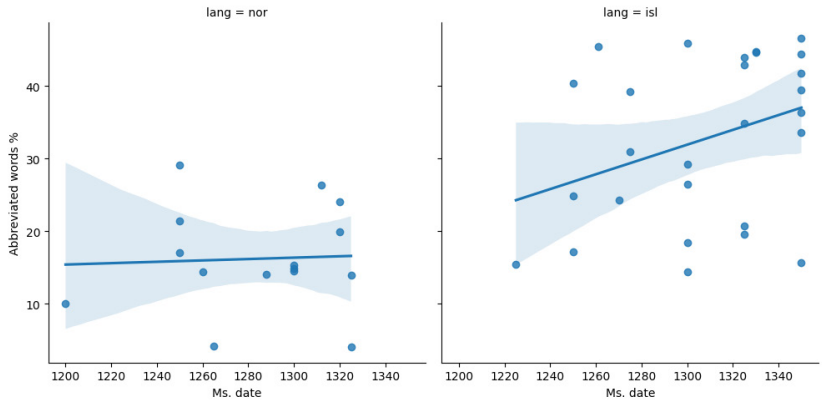

Figure 4: Abbreviated words in Menota Norwegian and Icelandic manuscripts before c. 1400.

In the Icelandic manuscripts there is a weak correlation (Pearson's $\rho = 0.35$) between the date and abbreviation rate of the available manuscripts before c. 1400. In the Norwegian manuscripts there is almost no correla-

tion ($\rho = 0.05$) between the date and abbreviation rate. Both diagrams in Figure 4 show that throughout the period before c. 1400, manuscripts are highly variable in their use of abbreviations, with both very low-abbreviating and very high-abbreviating manuscripts represented in all periods.

With respect to Hreinn's first observation in the introduction, the amount of abbreviation at the start of the period represented by the Menota archive (c. 1200) is similar to other vernaculars and less than for Latin manuscripts. Icelandic abbreviation never comes close to the extent of Latin abbreviation in its more extreme form, despite Hreinn's and others' claims. However, at the end of the period shown in the diagrams, there is much more significant divergence from the French and English vernacular manuscripts in both languages, with substantially more abbreviated words found in Icelandic and, to a lesser extent, Norwegian manuscripts than in other vernaculars, where abbreviation was slowly abandoned.

This data also confirms the second observation here by Hreinn, namely that Icelandic and Norwegian practice diverged and that abbreviation is more common in Icelandic manuscripts, but it does not fully support the observation that abbreviation increases over the period observed here, at least as applied to this corpus. Further data analysed below, however, strengthens this claim.

Extent of Abbreviation in Poetry and Prose

A few of the manuscripts in this category contain both poetry and prose, which allows us to compare the amount of abbreviation in the two categories. For the data shown in Table 4, words contained within <lg> elements (TEI line groups, used for poetry) are compared with all word tokens outside these elements. Certain manuscripts contain a very small amount of poetry and therefore insufficient data for this study. The manuscripts examined here therefore contain at least 1,000 poetic characters and both prose and poetry.

Ms	Date	Prose tokens	Prose economy %	Poetry Tokens	Poetry economy %
Digital editions from the Menota Archive					
AM 35 fol.	c. 1700	68560	18.8	5205	7.5
AM 63 fol.	c. 1700	94393	21.2	7763	8.1
AM 132 fol.	c. 1330–1370	114168	21.1	5672	8.0
AM 242 fol.	c. 1350	67339	8.2	13378	3.6
NRA norr fragm 52	c. 1225	3275	7.3	260	2.1
WolfAug 9 10 4to	c. 1330–1370	42453	20.4	2033	7.0
Digital edition from MenotaB / EAE					
GKS 1009 fol.	c. 1275	376562	22.8	30716	7.7

Table 4: Abbreviation economy of prose and poetry in Menota manuscripts containing both.

Supplementing the Menota XML is an additional manuscript, *Morkinskinna* (GKS 1009 fol.), whose data are taken from the MenotaB-based digital edition by Kjeldsen and imported into the *Editiones Arnamagnæanae Electronicae* (MenotaG-based) framework.[40] Kjeldsen's edition uses the same underlying data model as Menota and can therefore be confidently compared with the Menota data.

In these manuscripts the prose text is abbreviated between 2.3 times (AM 242 fol.) and 3 times (GKS 1009 fol., excluding NRA 52, which has very few tokens) more than the poetry, and this independent of the broad chronological spread of the manuscripts. AM 242 fol. is likely the manuscript in this list that uses the most space for the writing in it. As it is also the least abbreviated of the manuscripts here (apart from the early and fragmentary NRA 52, again an outlier), this further supports the notion that abbreviation is employed with the goal of economy of use of the writing surface. The corollary is seen in GKS 1009 fol. (*Morkinskinna*), which is probably the manuscript with the smallest writing.

40 Cf. https://eae.ku.dk and Wills, "Asynchronous Linked Editing."

Abbreviation of Poetic Text in Manuscripts

Skaldic stanzas are written in manuscripts often with a text that includes extrametrical pronouns and non-enclitic particles, which are unlikely to have been pronounced as syllables in the original metre. To take these features into consideration is difficult, however, because it would require a close alignment of the manuscript text with the reconstructed poetic text. While the Skaldic Project has digital versions of both, it would require a great deal of additional work (tokenising and aligning) to accomplish this. The Skaldic Project also has digital variants linked in some detail to the text, but it does not always record where the manuscript text has these additional metrical expansions – it only does so where there are additional variants, otherwise the removal of extrametrical pronouns and particles is not recorded as variants.

The Skaldic Project includes (at the time of writing) some 14,066 transcriptions of exactly 5,000 individual stanzas in 315 manuscripts (the overwhelming majority of which are Icelandic in provenance), around 2 million characters (excluding spaces and tagging) in total. All periods and types of manuscript are used, giving a very broad sample of the manuscript corpus as represented by manuscripts containing skaldic verse.

The main purpose of the transcriptions has been to aid editors in preparing their editions and reviewers in checking readings. While the transcriptions have not been reviewed and corrected to the same extent as the published editions, they have frequently been corrected by editors in the process of producing the editions. Where a particular transcriber's work has been deemed sufficiently inaccurate to mislead or confuse editors or the public, their transcriptions have been removed from the database and are not therefore included in this study.

The transcriptions are based on the traditions of diplomatic editing in Old Norse, where the abbreviations are expanded and represented in italics. For this we use the <i> element, which is used specifically and unambiguously in the project as the semantic equivalent of the <ex> element in TEI. This has the advantage that most HTML user agents (browsers) render idiomatic text as italic, consistent with the Old Norse diplomatic tradition.

The transcription guidelines for the Skaldic Project were distributed to editors and transcribers as part of the *Editors' Handbook*.[41] The transcribers who contributed this information include a range of editors and assistants, with the ten most prolific ones (in order of stanza transcriptions contributed and used here) being Valgerður Erna Þorvaldsdóttir, R. D. Fulk, Tarrin Wills, Emily Baynham, Katharina Seidel, Soffía Guðný Guðmundsdóttir, Hannah Burrows, Helen Appleton, Kate Heslop and Diana Whaley.

The data available for skaldic transcriptions should be comparable to the poetic data in Table 4: both contain poetic texts that have expansions of abbreviations marked up. Using the measure of abbreviation economy on the Skaldic Project transcription corpus allows us to compare this dataset with that of the Menota manuscripts above. The Skaldic Project's corpus may not include all the poetry that is recorded for a manuscript in the Menota corpus, however, but this should not affect our results greatly. The results of this comparison are shown in Table 5.

Siglum	Dating	Stanzas transcribed	Total chars	Unex	Econ. %	Menota poetry %
AM 35 fol	c.1675–1700	173	22751	20677	9.1	7.5
AM 63 fol	c. 1675–1700	109	15928	14244	10.6	8.1
AM 132 fol	c. 1330–1370	232	32023	29098	9.1	8.0
AM 242 fol	c. 1350	517	56762	55652	2.0	3.6
GKS 1009 fol	c. 1275	21	1982	1785	9.9	7.7
WolfAug 9 10 4°	c. 1330–1370	79	10793	9871	8.5	7.0

Table 5: Skaldic manuscripts compared with Menota manuscripts' poetry (see Table 4 above).

There is a small difference in the abbreviation economy of the poetry in the two datasets. The relative difference here is likely insignificant ($p = 0.1$ using a paired t-test) and in all but one case slightly lower in the Menota corpus than in the Skaldic Project corpus. This points to a slightly different expansion practice in the two corpora, which may be related to other differences such as tokenisation. In any case the relative amount of abbreviation in both corpora is highly consistent, with the ranking of each

41 Wills, *Editors' Manual*, 33–6.

manuscript by abbreviation economy being the same. We can therefore with confidence examine broader trends in abbreviation.

Figure 5 is a scatterplot of all manuscripts in the Skaldic Project's transcription data. Only manuscripts with at least thirty stanzas transcribed (128 in total) are included. Not all manuscripts in the database are marked as Norwegian/Icelandic, but of those that are represented here, only four are Norwegian. The horizontal axis represents the midpoint of the dating of the manuscript in the Skaldic Project's database and the vertical axis is the abbreviation economy as a percentage. The trendline in the graph is the locally weighted regression (LOWESS[42]), representing a smoothed overall trend.

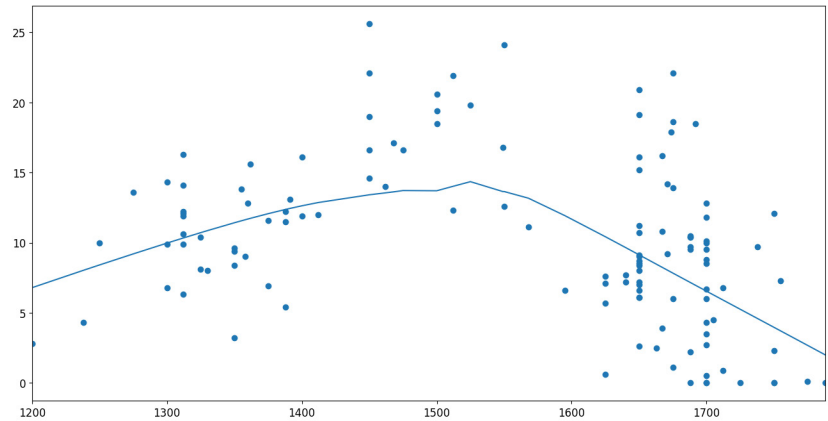

Figure 5: Abbreviation of skaldic stanzas in manuscripts from c. 1200–1800.

With this larger collection of manuscripts, clearer trends are observed than those in the Menota corpus, even though the corpus size itself is smaller. The relationship between manuscript date and abbreviation economy is more complex over this longer time period, increasing in the first centuries and then decreasing after the Middle Ages.

42 William S. Cleveland, "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association* 74, no. 368 (1979), 829–36; as implemented in the Python statsmodels module.

Period	Dates	Mss	Average economy %	Standard deviation	Min. econ. %	Max. econ. %
1	1200–1412	35	10.4	3.42	2.8	16.3
2	1450–1568	18	17.9	3.97	11.1	25.6
3	1595–1788	72	7.9	5.47	0.0	22.1

Table 6: Three periods of skaldic manuscript production.

Figure 5 suggests that abbreviation in a diachronic perspective falls into three periods, which can be described as follows (see Table 6):

1. Early manuscripts (prior to c. 1450): abbreviation rates increase over time with some variation, up to a maximum of c. 16%. In this period there is a clearer correlation between the manuscript date and increasing abbreviation usage (Pearson's $\rho = 0.6$, compared with $\rho = 0.35$ in the Menota Icelandic corpus).
2. Late medieval/Reformation manuscripts (between c. 1450 and the end of the sixteenth century): abbreviation is very consistently extensive, between 11 and 26%.
3. Post-Reformation manuscripts (from the end of the sixteenth century): abbreviation is overall lower than in the previous periods and decreases over time (Pearson's $\rho = -0.32$), but is highly varied (the standard deviation, a statistical measure of variance, here is 5.5, considerably higher than in the other two periods (3.4 and 4)).

The extensive variation of the third group may be due to a variety of reasons. It is possibly because of the divergence of purpose into two major types of manuscript writing after the Reformation: scholarly manuscripts that aimed to record accurately the palaeography and orthography of medieval manuscripts (and which we now often rely on where the original is lost) and 'lay' manuscripts which were copied for private and domestic purposes. The former could be expected to mirror medieval scribal practice, whereas the latter might reflect contemporary practices, even as they are written at the same time. The overwhelming majority of the manuscripts used by the Skaldic Project would fall into the first category, however, because the focus is on transcribing independent witnesses in that project. Another factor in this final period is the emergence of writing poetry in

lines rather than as inline prose. As skaldic metres generally consist of relatively short lines, there would be no advantage in saving page space by abbreviating poetry written in lines, as the space used remains the same. However, the Kringla manuscripts (AM 35 & 63 fol.) in Table 3, for example, lineate the poetry but show no significant difference in abbreviating prose relative to poetry compared with the other manuscripts there. A further investigation of lineation in skaldic manuscripts is required to understand this phenomenon.

The abbreviation economy for poetry is in general higher than that observed in the Menota corpus for poetry, but as we have shown before, when compared with the same manuscripts in the two corpora there is no significant difference. Given the correlation between abbreviation economy in prose and poetry, the above data would suggest that with a sufficiently large digital corpus of prose or prosimetric text marked up with expansions, we would observe a similar diachronic spread in the use of abbreviation.

Discussion

Returning to the four observations of Hreinn Benediktsson in the opening, we can largely confirm the observations he makes but with some refinements as regards Norwegian manuscripts, poetic and prose texts, and some further observations for the longer period of manuscript production in the Skaldic Project's transcriptions.

Regarding the first observation, that Icelandic and Norwegian medieval manuscript writing uses abbreviation more extensively than other vernaculars, this is true of Icelandic manuscripts, but less so of Norwegian ones. From the Menota data, abbreviation in Norwegian manuscripts (16%) appears in the early period to be consistent with that in Middle English manuscripts (around 10–20%[43]) and substantially lower than that observed in Latin manuscripts (up to 55%[44]). Icelandic manuscripts lie between the Latin manuscripts and other vernaculars, including Norwegian. Icelandic manuscripts diverge from both Norwegian and other vernaculars in that they increase their use of abbreviation towards the end of the Middle Ages,

43 Honkapohja and Liira, "Abbreviations and Standardisation," 282.
44 Ibid., 281.

whereas other traditions slowly abandon abbreviation during the same period. This suggests divergence from European practices, and reconvergence only occurs very slowly and late, long after the Reformation.

To this we can add the observation that there was a marked difference in the extent of abbreviation of poetry and prose, with poetry abbreviated much less than prose in the same manuscripts. There are only five or six manuscripts where poetry and prose can be reliably compared in the datasets used here, but these are so consistent that more data seems unlikely to alter this picture.

Icelandic manuscript writing uses more abbreviation than Norwegian: this is very much the case, starting with the very earliest period and increasingly diverging as the Middle Ages progressed, with a very significant difference in the period investigated here (c. 1200–1350) common to both traditions.

The observation that the amount of abbreviation increases in the first centuries of manuscript writing in Iceland is consistent with observations particularly of skaldic poetry transcriptions, but less so of the Menota corpus. To this we can add that at the end of the Middle Ages and into the early post-Reformation period, abbreviation remained very extensive in Icelandic manuscripts. After this period it began to be used much more sparingly, but with still considerable variation observed in this late period.

We cannot from these data determine the motivations behind abbreviation (i.e. whether it is driven by economy of page use), but it should be noted that the changes in abbreviation correlate with other developments in Iceland during the period studied. For example, the change in abbreviation economy in the first period identified above and observed in both the Menota and Skaldic Project transcription corpora correlates with the transition from the so-called Medieval Warm Period (to c. 1250) to the Little Ice Age (from c. 1450), where decreasing productivity of land may have put pressure on livestock production, in turn leading to a scarcity in parchment. The Black Death reached Iceland in 1402–4 and coincides also with the transition to the second phase of abbreviation practice identified here.[45] Conversely, the marked decrease in abbreviation economy

45 This observation was suggested by one of the anonymous reviewers of this paper. This event may also explain the gap after the start of the fifteenth century in relevant data from the Skaldic transcription corpus in Figure 5.

(Pearson's $\rho = -0.57$) from c. 1500 to c. 1800 coincides with the introduction of a cheaper material for manuscript writing, namely paper. The introduction of paper does not seem to coincide with a very sudden decrease in abbreviation, but scribal practices are likely to have taken time to adjust to the new technology.

The correlation between page material scarcity and cost on the one hand, and the economising of script by abbreviation, points to the hypothesis that abbreviation was driven by economic concerns in addition to orthographic trends. This in turn could suggest that poetry was of higher status, because it used more writing surface space than the same amount of text in prose. Either way, more resources were used relatively in writing poetry than prose, regardless of whether it is because it was seen as more valuable and therefore deserving of more resources, or more simply because it was necessary in order to record poetry more clearly. Another potential reason for why poetry is abbreviated less may be that it was considered more difficult for a potential reader to understand. The scribe may have therefore included more information about the text physically on the page, that is, removed less information by abbreviation. This would be consistent with the inclusion of extrametrical features often found in manuscript versions of poetry, which add extra information to aid in understanding the poetry.

Correlation does not, however, imply causation, and the changes in abbreviation usage coincided with a number of other shifts in writing practice. These phenomena could be investigated further, particularly: the economics of writing surfaces and a more absolute measure of writing surface use. The economics of producing writing surfaces requires a closer examination of livestock and parchment production as well as paper availability. Measuring writing surface usage would require actual measurement of the absolute physical space used by text rather than the relative measures shown here. Handwritten text recognition technologies and other spatial analysis systems such as MenotaG promise to make such studies possible in the near future.

The increase in abbreviation economy also raises a question about the materiality of text in Iceland in the course of the Middle Ages: in a sense, abbreviation represents the removal of increasing amounts of text from its material manifestation, and thus a kind of dematerialisation of text over

that period. In the earlier period, as is well documented in the Menota corpus, abbreviations are fairly standardised and can be understood by the reader because they have a limited number of potential expansions independent of context. In the central period, as represented by the Skaldic transcription corpus, abbreviation appears to be less determined, relying increasingly on the immaterial contexts of language and literature for the reader. This period is completely absent from the current Menota corpus of Icelandic and Norwegian manuscripts, and the Skaldic corpus does not provide unexpanded forms. With better data for this period, we could begin to understand the potential dematerialisation of text in Iceland.

BIBLIOGRAPHY

MANUSCRIPTS

(Several hundred manuscripts form the basis of the present study, but only those manuscripts that are referenced by siglum are included here.)

Stofnun Árna Magnússonar í íslenskum fræðum, Reykjavík

AM 132 fol.
AM 748 I b 4to
AM 757 a 4to

Den Arnamagnæanske Samling, Copenhagen

AM 35 fol.
AM 45 fol.
AM 63 fol.
AM 242 fol.

Royal Library, Copenhagen
GKS 1009 fol.

Riksarkivet, Oslo
NRA norr fragm 52

Herzog August Bibliothek, Wolfenbüttel
WolfAug 9 10 4to

SECONDARY SOURCES

Cleveland, William S. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74, no. 368 (1979): 829–36.

Cottereau-Gabillet, Emilie. "Revealing Some Structures and Rules of Book Production (France, Fourteenth and Fifteenth Centuries)." In *Ruling the Script in the Middle Ages. Formal Aspects of Written Communication (Books, Charters, and Inscriptions)*, ed. Sébastien Barret, Dominique Stutzmann, and Georg Vogeler, 129–63. Utrecht Studies in Medieval Literacy 35. Turnhout: Brepols, 2016. DOI: http://doi.org/10.1484/M.USML-EB.5.112434.

Dahlerup, Verner, ed. *Ágrip af Noregs konunga sögum*, STUAGNL 3. Copenhagen: S. L. Møller, 1880.

Driscoll, Matthew James. "Marking up Abbreviations in Old Norse-Icelandic Manuscripts." In *Medieval Texts – Contemporary Media: The Art and Science of Editing in the Digital Age*, ed. Maria Grazia Saibene and Marina Buzzoni, 13–34. Pavia: Ibis, 2009.

Gottskálk Jensson. "Udgivelse af norrøn litteratur indtil 1772." *Dansk Editionshistorie 2: Udgivelse af norrøn og gammeldansk litteratur*, ed. Britta Olrik Frederiksen, 47–106. Copenhagen: Museum Tusculanum, 2021.

Haraldur Bernharðsson and Odd Einar Haugen. "Chapter 6. Abbreviations." *Menota Handbook*, ed. Odd Einar Haugen, 3rd ed. Bergen: Medieval Nordic Text Archive, 2019. https://www.menota.org/HB3_ch6.xml.

Haugen, Odd Einar. "The development of Latin script I: In Norway." In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages* 1, ed. Oskar Bandle et al., 824–32. Berlin & New York: Walter de Gruyter, 2002.

Haugen, Odd Einar (ed.). The Menota Handbook: Guidelines for the Electronic Encoding of Medieval Nordic Primary Sources, Version 3.0. Bergen: Medieval Nordic Text Archive, 2019. http://www.menota.org/handbook.xml.

Honkapohja, Alpo. "'Latin in Recipes?' A Corpus Approach to Scribal Abbreviations in 15th-Century Medical Manuscripts." In *Multilingual Practices in Language History: English and beyond*, ed. Päivi Pahta, Janne Skaffari and Laura Wright, 243–71. Berlin: De Gruyter, 2018. DOI: http://doi.org/10.1515/9781501504945-012.

Honkapohja, Alpo and Aino Liira. "Abbreviations and Standardisation in the Polychronicon: Latin to English, and Manuscript to Print." In *The Multilingual Origins of Standard English (MOSTE)*, ed. Laura Wright, 269–316. Berlin: De Gruyter. DOI: http://doi.org/10.1515/9783110687545-010.

Honkapohja, Alpo. "Digital Approaches to Manuscript Abbreviations: Where Are We at the Beginning of the 2020s?" *Digital Medievalist* 14 (2021) DOI: http://doi.org/10.16995/dm.88 https://journal.digitalmedievalist.org/article/id/7025/.

Hreinn Benediktsson. *Early Icelandic Script as Illustrated in Vernacular Texts from the Twelfth and Thirteenth Centuries*. Íslenzk handrit/Icelandic Manuscript Series in Folio II. Reykjavík: The Manuscript Institute of Iceland, 1965.

Kjeldsen, Alex Speed. *Filologiske studier i kongesagahåndskriftet Morkinskinna*. Bibliotheca Arnamagnaeana. Supplementum 8. Copenhagen: Museum Tusculanum, 2011.

Kölbing, Eugen, ed. *Elis saga ok Rosamundu*. Heilbronn: Henninger, 1881.

Kålund, Kristian. *Palæografisk Atlas. Ny serie. Oldnorsk-islandske skriftprøver c. 1300–1700.* København: Gyldendal, 1907.

Pearson, Karl, "Notes on Regression and Inheritance in the Case of Two Parents." *Proceedings of the Royal Society of London* 58 (1895): 240–2.

af Petersens, Carl, ed. *Jómsvíkinga saga (efter Cod. AM. 510, 4:to) samt Jómsvíkinga drápa*. Lund: C. W. K. Gleerup, 1879.

af Petersens, Carl, ed. *Jómsvíkinga saga efter Arnamagnæanska handskriften N:o 291. 4:to*, STUAGNL 7. Copenhagen: S. L. Møller, 1882.

Seip, Didrik Arup. *Palæografi. B: Norge og Island*. Nordisk Kultur 28:B. Stockholm: Albert Bonnier, 1954.

Sperberg-McQueen, C. M. and Lou Burnard, eds. *Guidelines for Electronic Text Encoding and Interchange*. Chicago and Oxford: TEI P3 Text Encoding Initiative, 1994.

Stefán Karlsson. "The Development of Latin Script 2: In Iceland." In *The Nordic Languages. An International Handbook of the History of the North Germanic Languages* 1, ed. Oskar Bandle et al., 832–40. Berlin & New York: Walter de Gruyter, 2002.

Petti, Anthony G. *English Literary Hands from Chaucer to Dryden*. London: Arnold, 1977.

TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, Version 4.7.0. (TEI Consortium, 16 November 2023) http://www.tei-c.org/Guidelines/P5/ (accessed 23 March 2024).

Wills, Tarrin. The Foundation of Grammar: An Edition of the First Section of Óláfr Þórðarson's Grammatical Treatise, PhD Thesis. University of Sydney, 2001.

Wills, Tarrin, Diana Whaley, Kari Gade, Margaret Clunies Ross, Edith Marold, Guðrún Nordal, Valgerður Erna Þorvaldsdóttir and Kate Heslop. *Skaldic Poetry of the Scandinavian Middle Ages: Editors' Manual*. 2nd ed. Sydney: Centre for Medieval Studies, University of Sydney, 2005.

Wills, Tarrin. "Asynchronous Linked Editing of Texts in Physical Objects." Digital Humanities in the Nordic and Baltic Countries, Reykjavík 27–31 May 2024 (DHNB 2024, forthcoming).

DATASETS

Medieval Nordic Text Archive (Menota): Catalogue https://clarino.uib.no/menota/catalogue (2 March 2024).

Medieval Unicode Font Initiative: Combining Characters https://mufi.info/q?p=mufi/browse/tag/2 (9 July 2024).

MenotaG: Relative punctuation width data https://menotag.ku.dk/q?q=menotagdatapunctsize (23 March 2024).

Skaldic Poetry of the Scandinavian Middle Ages (Skaldic Project): All transcriptions https://skaldic.ku.dk/q?q=skpalltranscriptions (3 March 2024).

SUMMARY

Abbreviation in Old Norse manuscripts — *a quantitative study*

Keywords: Old Norse, vernacular manuscripts, abbreviation practices, diachronic analysis, manuscript studies

Previous scholarship on the amount and distribution of abbreviation in manuscripts has noted that Icelandic manuscripts use more abbreviation than other vernaculars and that this increases in the medieval period. This study investigates these and related observations *quantitatively,* using the editions and transcriptions of the Menota and Skaldic projects, and refines them with respect to poetry and prose, and compares them with new studies on abbreviation in Latin and vernacular manuscripts. It is observed that the extent of abbreviation in Icelandic and Norwegian manuscripts relative to other traditions may have been overstated, but that Icelandic manuscripts in particular diverge from other traditions increasingly over time. A substantial difference is further observed in the abbreviation of poetry and prose in manuscripts that combine them, with the prose normally abbreviated around three times as much as poetry.

This paper also develops a new measure of abbreviation based on marked-up expansions, showing the amount of writing surface area saved (abbreviation economy). This measure is closely comparable to the main existing measure in scholarship (proportion of abbreviated words) but can be applied to un-tokenised digital texts which only have expansions marked up. This measure is then applied to the Skaldic Project's transcription database. The results give a long-term diachronic perspective on abbreviation, showing that abbreviation economy can be divided into three distinct periods, rising in the course of the Middle Ages, remaining extensive through the Reformation and then gradually declining up to the start of the nineteenth century.

ÁGRIP

Styttingar og bönd í norrænum handritum — megindleg rannsókn

Efnisorð: norræn fræði, handrit á móðurmáli, styttingar og bönd, söguleg greining, handritafræði

Fyrri rannsóknir á fjölda og dreifingu á notkun styttinga og banda í handritum hafa bent til þess að í íslenskum handritum séu skammstafanir meira notaðar en í öðrum löndum og notkun þeirra fari vaxandi á miðöldum. Í þessari rannsókn er sjónum beint að fyrri athugunum og notaðar megindlegar aðferðir þar sem litið er á útgáfur og uppskriftir handrita úr Menota-textasafninu og dróttkvæðaverkefninu (Skaldic project). Sérstaklega er horft til munar á bundnu og óbundnu máli, sem og nýrra rannsókna á styttingum í handritum rituðum bæði á latínu og á öðrum málum. Athuganirnar leiða í ljós að umfang styttinga og banda í íslenskum og norskum handritum samanborið við aðrar hefðir gæti hafa verið ofmetið en að með tímanum hafi íslensk handrit skorið sig úr frá því sem tíðkaðist annars staðar. Verulegur munur sést einnig á styttingum bundins máls og óbundins í handritum sem hafa hvort tveggja þar sem lausamálstexti er yfirleitt styttur um það bil þrisvar sinnum meira en texti í bundnu máli.

Í þessari grein er einnig gerð grein fyrir þróun á nýrri mæliaðferð fyrir notkun á skammstöfunum sem byggist á gögnum úr mörkuðum textaútgáfum og leitt hefur í ljós hversu mikið pláss sparast með notkun þeirra (styttingarhagkvæmni). Þessi mæliaðferð er sambærileg við helstu núverandi mælingar sem tíðkast í fræðunum (hlutfall skammstafaðra orða) en þó er hægt að beita henni á ómarkaðan stafrænan texta þar sem eingöngu er gefið til kynna að leyst hafi verið upp úr böndum og styttingum. Þessari mæliaðferð er síðan beitt á gagnagrunn dróttkvæðaverkefnisins. Niðurstöðurnar gefa skýrar vísbendingar um notkun skammstafana yfir lengri tíma og sýna að henni megi skipta í þrjú ólík tímabil: hún fer vaxandi á miðöldum, heldur áfram að vera umfangsmikil fram yfir siðaskiptin og fer síðan smám saman minnkandi fram á byrjun nítjándu aldar.

Tarrin Wills
Njalsgade 136
2300 København K
Denmark
tarrin@hum.ku.dk