



Open Library of Humanities



Part of the Ubiquity
Partner Network



Review Article

How to Cite: Honkapohja, Alpo. 2021. "Digital Approaches to Manuscript Abbreviations: Where Are We at the Beginning of the 2020s?" *Digital Medievalist* 14(1): 1, pp. 1–40. DOI: <https://doi.org/10.16995/dm.88>

Published: 09 July 2021

Peer Review:

This is a peer-reviewed article in *Digital Medievalist*, a journal published by the Open Library of Humanities.

Copyright:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Open Access:

Digital Medievalist is a peer-reviewed open access journal.

Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

REVIEW ARTICLE

Digital Approaches to Manuscript Abbreviations: Where Are We at the Beginning of the 2020s?

Alpo Honkapohja

University of Edinburgh, GB
alpo.honkapohja@gmail.com

Abbreviations have been an important qualitative means for dating and localising manuscripts. In digital scholarship, however, they have received less attention. Reasons for this range from digital resources inheriting editorial traditions from print to normalisation being a prerequisite for many research questions. The aim of this paper is to build bridges by giving an overview of scholarship into digital and quantitative approaches – taking into account English, French, Old Norse and, to a lesser extent, Dutch, German and Celtic scholarship. It also makes a theoretical contribution by placing abbreviations into a typology of writing systems and proposing that the terms conditioned and unconditioned variation in analogy with phonology could be useful for studying abbreviation.

Keywords: manuscript abbreviations; digital palaeography; quantitative palaeography; manuscript studies

1 Introduction

§1 In her entry for the *Encyclopedia of Language and Linguistics*, Lowe defines the aims of palaeography as follows:

The basic task of palaeography is to provide the means of dating and localizing manuscripts by establishing patterns in the development of characteristic letter forms and abbreviations. (Lowe 2006: 134)

As the entry recognises, abbreviations have been one of the most important means for dating and localising manuscripts available to palaeographers and book historians. In digital scholarship, however, they have been something of a sidetrack. Reasons range

from many digital resources inheriting editorial traditions from print to the labour-intensiveness of rich XML transcription and normalisation being a prerequisite for many research questions. Moreover, to some extent, the focus of digital scholarship simply was elsewhere in the 2010s. Much work concentrated on achieving efficient results via lemmatisation, alignment, and digitisation. Furthermore, recent research in quantitative palaeography (e.g. Kwakkel 2012; Stokes 2014; Thaisen 2017) often focused on characteristic letterforms rather than abbreviations. Nevertheless, the importance of abbreviations has been acknowledged by many scholars (e.g., Hasenohr 2002; Driscoll 2009; Stutzmann 2014; Kestemont 2015). There have also been a number of quantitative studies since the 1980s that have uncovered interesting results, which this review seeks to highlight.

§2 The aim is to give an overview of scholarship into digital and quantitative approaches – taking into account English, French and German and, to a lesser extent, Dutch, Old Norse and Celtic scholarship. My hope is that the review will help anyone interested in the subject to find secondary literature and to identify trends and common interests. As many of these findings are in languages other than English, one motivation is also to help their dissemination into Anglophone scholarship. Furthermore, I hope to make a contribution to the theoretical description of abbreviations by placing them in the context of a typology of writing systems, drawing on writing systems research, which has been emerging as an important new field with contributions, such as Daniels and Bright (1996) and Cook and Ryan (2016).

2 Recognising the value of trivial or accidental variation

§3 While the importance of abbreviations is acknowledged by some scholarship, there are many other fields where there is a long tradition of treating them as a problem rather than evidence. The chief culprits are editorial practices and textual scholarship aimed to restore texts to their original form (Driscoll 2010). The twentieth-century editorial theory often focused on authorial “work” rather than scribal “text,” treating abbreviations and other scribal variation as “accidentals” (see e.g. Greg 1951), which were not seen as relevant for the authorial work contained in the manuscript copies. Much scholarship focused on the work and uncovering the work under layers of scribal copying and errors. Even when editions are based on a

single manuscript witness, they very often contain tacitly expanded abbreviations, as they were considered to be a barrier to reading and understanding the text. The messiness of “accidental” variation was thus something which the editor attempted to get rid of. It has also been a problem for the requirements of digital scholarship.

§4 Text retrieval systems are typically unable to recognise different forms of the same word and the problem is usually solved by normalisation (cf. Kestemont 2015, 160). Many research questions are unlikely to be successful “on collections of texts that do not adopt the same orthographic standard” (161). These include topic modelling, computational semantics, text mining, and stemmatology as well as some fields of corpus linguistics such as syntax. Moreover, some of the research areas and questions that evoked most interest in pre-digital scholarship have been among the first to receive digital equivalents. For example, digital stemmatology inherits the outlook of more traditional approaches in which the focus is on features that are likely to be part of the “work.” However, some have questioned this, including Andrews and Mace who call into question “the utility of excluding “trivial” variation such as orthographic and spelling changes from stemmatic analysis” (2013, 504). Moreover, digital corpora are often based on printed editions and inherit their editorial practices, sometimes conflating several editorial approaches in the same resource (Honkapohja, Kaislaniemi and Marttila 2009, 456–460). In short, for several reasons, normalisation and the expansion of abbreviations are very much the norm in digital scholarship.

§5 The main problem with normalisation is, however, that while it is necessary for some research questions, it also discards a large amount of potentially useful data, which makes other types of research impossible. As Kestemont (2015, 161) puts it: “superficial textual variations also present important scholarly opportunities, for instance for the identification of scribes or the dialectological analysis of texts.” The problems caused by the loss of data, resulting from discarding variation considered accidental or trivial, has been discussed and criticised by several scholars (Driscoll 2006; Kytö et al. 2011; Kopaczyk 2011; Rogos 2011; 2012; Stutzman 2014; Lass 2004). This criticism is supported by the fact that there are numerous approaches which have demonstrated the value of scribal and other accidental variation.

§6 Making better use of “accidentals” or “trivial variation” is central to quantitative approaches to palaeography. It can also be linked to what can be termed as “new philology” or “material philology” (cf. Driscoll 2010). Material philology has been calling attention to recognising the uniqueness of each manuscript copy, including codicological features, such as writing support and level of decoration, but also all of the accidentals, such as punctuation, spelling variation, lexical variation as well as abbreviations. All of these are related: “orthography, palaeography, and codicology are overlapping realms of the archaeology of the book” (Thaisen 2011, 84). Nevertheless, there are decisions that need to be made on how much of them to take into account.

§7 The decisions are on which level to focus when studying variation. The focus of this review article is mainly on the graphemic level, as my concern is with different types of signs and their referents, not “allographic” variation between various shapes of the same sign (see e.g. Robinson and Solopova 1993, 20). In making this decision, my focus differs somewhat from much recent work, which has opted for differences on the “graphetic” level, recording allographic variation (ORIFLAMMS project, Rogos 2011; they also include studies of variant letter forms such as Thaisen 2011; Stokes 2014; Kwakkel 2012; Speed Kjeldsen 2013). The reason for this is that even though abbreviations are connected both to graphetic variation and the codes on the page, they are also part of medieval writing systems and made their orthography more complex than alphabetic writing. This complexity is something that deserves to be addressed as an important phenomenon in itself.

3 Abbreviations in the typology of writing systems

§8 What makes abbreviations unique, compared to many other features deemed accidentals by textual criticism, is the range in the way the written form corresponds to its lexical and phonological referents. Abbreviations belong to the “grey area” components of the medieval manuscript – which Traube called *Halbgraphische Objekte* (“half-graphic objects”) (Benskin 1977, 506; Römer 1993 and 1997; Rogos-Hebda 2018, 58). However, while the non-alphabetic nature of abbreviation is noted by several scholars (LAEME 2.2.1.; see also Benskin 1977, 506; Rogos 2011, 47; 2012,

7), it is often discussed in somewhat imprecise terms or with much variation in terminology. This can be partly explained by different emphases, as some approaches are concerned with allographic variation, codicological phenomena or morpho-phonemic variation. (For example, Mazziotta [2008] aims to describe abbreviations from a graphetic perspective seeking to model the strokes made by the scribe. Rogos [2011, 2012], Rogos-Hebda [2018], Cottureau [2005] and Cottureau-Gabillet [2016] discuss abbreviation in addition to range of other bibliographic codes such as initials. LAEME[2013] is considered with phonology and morphosyntactic description.)

§9 The situation is not helped by using different names for the same phenomena in different fields of scholarship. As Römer (1997) notes, where linguists may speak of *apocope*, palaeographers speak of *suspension* (11). If we, however, focus precisely on the signs and practices used to abbreviate words in mediaeval writings and their lexical and phonological referents, it is also possible to describe them in terms of writing systems research and terminology developed by scholarship describing the emergence of writing systems in the ancient Middle East and shores of the Mediterranean.

§10 “The main issue” with classifications of writing systems “is how to reconcile the two levels of language that written symbols correspond to [...] the lexicon, whether words or morphemes [...], and on the other hand sounds in the phonology, whether syllables (Japanese kana), phonemes (Italian) or consonant phonemes only (Arabic)” (Cook 2016, 6). The following account uses terminology used in a pioneering monograph by the Polish-born Assyriologist I.J. Gelb (1952), which aimed to describe the development of writing systems from a typologically comprehensive perspective (for a very comprehensive resource, see Daniels and Bright 1996; for a clear, if somewhat contentious, introduction, see Powell 2009).¹ The study introduced

¹ The comparative study of writing systems is also a field that has somewhat suffered from a large amount of variation in terminology – this is especially since it is studied by scholars working in fields as diverse as palaeography, epigraphy, linguistics, philology but also archaeology and Egyptology. “What we have, instead, are narrow fields of study of the type of Semitic epigraphy, Arabic paleography, Greek or/and Latin epigraphy or/and palaeography, Chinese palaeography, papyrology, etc., all limited to certain periods and geographic areas. In all cases these narrow fields of study form

a “tripartite scheme of logography, syllabary, alphabet” which has become the main typology used in writing systems research (Daniels and Bright 1996, 8). Abbreviations too can be described by this framework, which Gelb (1952, 12–13) already notes. (Gelb’s [1952] example, the following sentence contains examples of all three: “Mr. Theodore Foxe, age 70, died to-day at the Grand Xing Station” [13]. As Gelb notes both the numeral “70” and the abbreviation Mr. for “Mister” are logograms, whereas “the rebus-type symbol X plus alphabetic ing” stand for the word “crossing” [ibid.]). The typology works well, if one takes it as a somewhat flexible system and does not expect scripts to fall neatly under any category (Daniels and Bright 1996, 8). The placement of abbreviation symbols in this typology is illustrated by **Figure 1**.

§11 Writing systems can be divided into *phonography*, in which signs are tied to sounds, *logography*, in which signs are tied to words, and *semasiography*, which is a broad umbrella term for symbolic systems that communicate meaning without being directly tied to natural language.² Phonographic writing systems can be further divided into *syllabography* and *alphabetic writing*. Syllabic systems refer to the basic building blocks of spoken utterance: syllables. They include Japanese *kana* or Korean *hangul*. Alphabetic systems link to vowel and consonant sounds; “a writing system

subdivisions of wider, but still specific, fields of study, such as Semitic or Arabic philology, classical philology, Assyriology, Sinology, and Egyptology” (Gelb 1952, 22).

2 I am not entirely satisfied with the term *semasiography*, but alternatives, such as *pictogram*, *icon* or *symbol*, are too narrow, broad or come with additional theoretical baggage. Perhaps the most obvious term might be “pictogram,” but unfortunately, as Powell (2009) notes, the term has been used so sloppily – it can refer to logograms like Egyptian hieroglyphs or to semasiograms – it is practically useless as theoretical distinction in serious scholarship. “Icon,” on the other hand, has a wide range of meanings from religious art to graphical user interfaces. In linguistics, iconicity refers to similarity or analogy between a sign and its referent. This renders it partly incompatible with *semasiography*, as it may also include uses in which the relation between a sign and its referent is conventional and non-iconic. The term “symbol” has even broader range of uses and definitions than icon. The term is used in various fields, including but not limited to semiotics, psychoanalysis and literary scholarship. This makes it too broad for the purposes of the distinction intended here, between writing systems tied to natural language and ones that are not. Consequently, I prefer *semasiography* as an umbrella term for writing in which the signs are not tied to necessary parts of speech. The term may be unfamiliar for many, but it is more precise than pictogram and lacks the theoretical baggage of icon and symbol. It can also be taken as a tribute to I. J. Gelb, whose pioneering account section 3 is builds on. Moreover, it provides a link with semiotics, a field that was very popular in the late 20th century, but has recently fallen out of fashion, but could provide useful subcategorisations within the very broad field of *semasiography*.

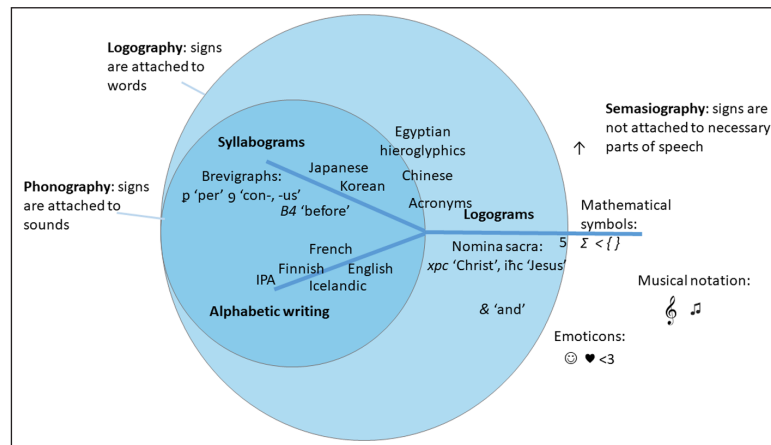


Figure 1: Typology of writing systems.

in which each symbol corresponds to a particular sound of the language, and, vice versa, each sound corresponds to a symbol, is called ‘transparent’ or ‘shallow’ (Cook 2016, 7). A system like the International Phonetic Alphabet (IPA) gets very close to complete transparency, and smaller European languages like Finnish or Icelandic have fairly shallow orthographies, but the spelling systems of languages like English and French are less transparent. For example, English orthography carries such historical baggage as the “silent” final “gh” in *eight*, *night*, *through* and there are homophones such as *dear/deer* –in which, Laing argues, the distinction indicated by spelling convention is leaning towards logography (LAEME 2.2.1). In logography, signs are attached to entire words. Chinese script or Egyptian hieroglyphics are more logographic, even though both have phonographic connections (Powell 2009, 188). It is, however, also possible for written symbols to communicate meaning without having a direct phonetic or lexical referent.

§12 I call writing systems which are not connected to natural language semasiographic, following Gelb (1952) and Powell (2009). Semasiography can be defined as “material marks with a conventional reference” that “communicate information without the necessary intercession of forms of speech” (Powell 2009, 32). Semasiography is a broad and heterogeneous category which includes systems that predate phonographic and logographic writing, but also still exist beside them. Modern semasiographic signs include, for instance, emoticons, computer icons and

traffic signs. Semasiographic systems can be complicated and precise, as they also include, for example, mathematical and musical notation. Yet, it would be impossible to write this article in mathematical or musical notation without giving the numbers or notes some kind of phonographic or logographic referents.

§13 Western scripts are predominantly alphabetic, but they also routinely contain characters that are syllabographic or logographic. An important point about the classification, which Gelb (1952) already notes is that the signs can be used flexibly depending on the context (16–17). For example, the heart shape (♥ or <3>), which can carry meanings such as indicating one of the four suites in a standard 52-card deck (**Figure 2**) or communicating love or other strong emotion as an emoticon, falls under semasiography. In examples 1 and 2, it is however used as a logogram, but referring to different words, which most English speakers have no problem parsing as “love” and “heart.”



Figure 2: Semasiography (Wikimedia Commons, 2020).

1. I ♥ NY
2. My ♥ belongs 2 U.

What is more, example 2 contains two modern abbreviations categorised as letter/number homophones by Bieswanger (2007, 5), which are used outside their normal category (see also Anis 2007). The Hindu-Arabic numeral 2 is normally a logogram, but in the example, it is used as a syllabogram due to phonological similarity /t^hu:/ with the English, single-syllable, preposition *to* /tu/. A second example of the same is the letter “U,” one of the letters of the Latin alphabet, but here serving as a syllabogram as its name, when read out aloud in English is homophonous with the single-syllable second person pronoun “you” /ju:/. The fact that these are syllabograms rather than logograms is evident, as you need two signs to write a two-syllable word such as “before”: *B4*. It is thus very much possible to use this system to describe abbreviation: modern as well as medieval.

§14 Medieval abbreviations can be easily described using this same typology. Some abbreviations (° “-er, -ir, -re” in *af^o*) are syllabograms consisting of vowel and consonant combinations (often “s” or “r” in combination with a vowel). Others (*S.* for “Saint” – or indeed modern acronyms) are logograms, as they correspond to the entire word. Within logograms, Laing makes a further distinction between “impure” logograms, which contain some phonographic cue, such as *þ* “that,” *S[aint]*, writing common religious phrases with initials only (LAEME 2.2.1). On the other hand, Laing considers the use of Greek abbreviations in Latin such as *xpc* for “Christ” and *iħc* for “Jesus,” or the ampersand, to be pure logograms, as they are tending towards abstraction as they are “not subject to phonological extrapolation” (LAEME 2.2.1). Even if they are made up of letters of the Greek alphabet and were often reproduced by the scribes using more familiar Latin shapes, they do not have these values (LAEME 2.2.1).

§15 An important point to make about abbreviations, though, is that even though they may function like syllabograms or logograms, they are additional and alternative signs to an alphabetic system, not a new form of syllabic or logographic script to represent spoken utterances. Kopaczyk (2011, 96) notes that abbreviations

should be treated as a sequence of letters rather than syllables, which is evident from examples such as this *co[ur]t* or *p[er]sone* (Kopaczyk 2011, 96, citing an example in Roberts 2005, 11). Driscoll (2006, 14) gives another example, *Pīg* “*pi[n]g*.” The scribe appears to have been thinking of the nasal sound [ŋ] as consisting of the digraph “ng” like it is represented in the Latin alphabet. Rather than a new syllabic script, abbreviations are thus an elaboration of the Latin alphabet. They belong with other auxiliary devices, such as diacritics, capitalisation and punctuation.

§16 Two more useful pieces of terminology are relevant for describing abbreviations as auxiliary devices in writing systems. Powell (2009, 42) makes a distinction between *phonetic complements* and *semantic complements*. An example of a phonetic complement would be the addition of letters, such as 2nd to indicate the ordinal “second.” In the medieval system, superscript abbreviations (*w* “with”) typically function as phonetic complements. Semantic complements or determinatives, on the other hand, specify the meaning of a sign with an additional mark. Latin letters such as a bar through the ascender of a *p* “crossed-p” are semantic complements indicating that the letter stands for the abbreviation for “per, par, por” in Latin or English. The term semantic complement is also useful for those mediaeval abbreviations which Cappelli (1990 [1899], xxiii–xli) calls *abbreviation marks significant in context* rather than *abbreviation marks significant in themselves*. (Cappelli’s terms are in Italian *Segni abbreviativi con significato proprio* [xxiii] and *Segni abbreviativi con significato relative* [xxix]. I am using the English translations by Heimann and Kay [1982].) Some medieval abbreviations, such as *punctus* or the horizontal bar based above the abbreviated word, simply indicate the presence of an abbreviation. These signs are best described as semantic complements rather than logograms or syllabograms.

3.1. Problems related to expanding syllabograms and logograms

§17 Some of the problems related to expansion of abbreviations, practised in many fields, are caused by alphabetising texts written in a writing system whose orthography allowed also non-alphabetic characters. Orthography can be defined as “the rules for using a script in a particular writing system, that is to say how the

symbols spell out words etc” (Cook 2016, 6). The orthography of Medieval Latin and vernaculars permitted the use of a sometimes large number of syllabograms and logograms – the number of Latin abbreviations can be more than 50 per cent of the word count (see, e.g., Honkapohja and Liira, 2020; Stutzmann 2018); in vernacular English or French the number may be as high as 30 per cent. In Old Norse, the number could exceed even mediaeval Latin (Driscoll 2009, 13).

§18 The inherent difficulty with expanding abbreviations lies in the fact that logograms and syllabograms are less phonetically transparent than alphabetic script. A mediaeval text which contains logographic and syllabographic characters mixes writing systems. Such a system can, following Mazziotta (2008, §2.1.1), be called *plurisystème*, a “multi-system.” (Even though, strictly speaking, Mazziotta was referring to his own scribe-oriented palaeographical model of abbreviations). If the abbreviations are expanded and represented by characters of Latin script, a modern editor assigns a definite alphabetic and implied phonetic value to them (Rogos 2012, 7). The edited text becomes a mixture of scribal and editorial language. The problem is greater with medieval vernaculars, whose spelling was notoriously varied, in contrast to the relatively homogenous spelling of mediaeval Latin.

§19 The usual solution is to use the scribe’s most common unabbreviated form as the expansion (cf. Page 1960: xxxii; De la Cruz-Cabanillas and Diego Rodriguez 2018). However, this may seriously distort the frequencies. For example, Smith (2018: 192, Table 9.1.) found in her Older Scots data that the scribes abbreviate the plural of nouns with *ʃ* 61 per cent of time (e.g., *partʃ* “parts”), followed by “-is” (20%), “-es” (11%), “-ys” (5%) and “-s” (5%). This is important as the practice of spelling the plural with “-is” or “-ys” (*partis* or *partys*) instead of “-es” (*partes*) is considered diagnostic of Older Scots in contrast to more southern dialects. An editorial approach that expands all abbreviated plurals following the most common spelling will lead to showing “-is” as the clearly dominant variant with 81%, when in reality there is no certainty which form the scribe intended (see also Czajkowski 2018, 96).

§20 De la Cruz-Cabanillas and Diego-Rodriguez (2018, 172), on the other hand, give a few Middle English examples of words whose expansion may have dialectal significance for the tradition of Middle English dialectology in which spelling

variation is used to localise texts. For example, the word *man*, if abbreviated with a bar, may be interpreted as otiose, but may also indicate abbreviation. The form *man̄* can be expanded either as “man,” “mane,” “mann” or even “maun” (if the two minims are interpreted as a “u”). (All of these forms are recorded by the *Linguistic Atlas of Late Mediaeval English* [eLALME, 2013 (1986)]). The expansion matters, because *man* is a very widely used form, whereas *mane* or *mann* are much more restricted and can be used as diagnostic forms (cf. Cruz-Cabanillas and Diego-Rodriguez 2018, 172). Consequently, there is often no certainty of how a certain abbreviation should be expanded – mixing writing systems that function on different levels can lead to problems like these.

3.2. Logography and language independence

§21 A further feature of logograms is that they are less tied to a single language than phonographic writing. For example, the Hindu-Arabic numeral 2 can be read out aloud in any language. An English person would expand it as “two,” a French-speaker as “deux” and a Finnish speaker as “kaksi.” Abbreviations, including less straightforwardly logographic ones, can sometimes too be expanded in several different languages. As Voigts (1989, 91) mentions, “[n]o educated reader is perplexed by *e.g.*, *i.e.*, or even *viz.*, but it is by no means certain that even the latinate individual actually thinks *exempli gratia*, *id est*, or *videlicet* when he reads or writes those letters.” There is some contemporary evidence that scribes could associate Latin abbreviations with the vernacular. The following examples are from three closely related copies of a plague treatise in medical manuscripts from the 1450s and 1460s (Honkapohja 2017, 136). The scribe of Sloane 3566 writes English “that is to say” where the other scribes use the Latin abbreviation *s.* for “scilicet.”

- The hede veyne lyeth aboue þe body veyne **s.** cardiac. (Trinity, f. 80v)
- The hede veyne lieth aboue the bodye veyne **s.** Cardiak. (Boston, f. 48v)
- The hede vayne lyeth a bove þe body vayne **þat is to sey** þe cardiake. (Sloane 3566, f. 97v)

[The head vein lies above the body vein, namely, cardiac]

§22 A number of studies have uncovered evidence that this language-independent quality was sometimes utilised in historical texts. Hector (1958, 37) mentions that English proper names in Latin documents could be “terminated by a mark of suspension to preserve the fiction that they were declinable Latin words.” It has even been argued that the language-independent quality of abbreviations may have been used on purpose.

§23 Abbreviated words that can be expanded in several languages are called *visual diamorphs* by Ter Horst and Stam (2018, 234),³ who focus on Latin and Gaelic. According to them, “[s]ome words are abbreviated ambiguously, and they can consequently be resolved as both Latin and Irish. One example is the title *aps.*, which can stand for “apostle” in Latin, *apostolus*, or Irish: *apstal*” (Ter Horst and Stam 2018, 234). Wright has studied abbreviations in English/Anglo-Norman/French mixed-language documents (see e.g. Wright 2002, 2011 and 2018). The author stresses the importance of abbreviations in the complicated contact situation during the late Middle English period, suggesting that the abbreviation system may be part of the reason, as it can be used to suppress the language-specific grammatical endings and highlight the stem. For example, a word such as *argent*⁹ can be read as Latin (e.g. *argentem*, *argentis*), Anglo-Norman French (*argenté*), or Middle English *argent* “(a) Silver, silver coin; (b) *her*. Silver-coloured, silver-gilt” (MED). Czajkowski (2018, 90) notes that abbreviated forms of pronouns can be used to suppress differences between High German and Low German forms of personal pronouns, as they can be expanded both to Low German “he,” “we” and “unse” or High German “er,” “wir” and “unsere.” Consequently, abbreviations have an important function in language-independent communication due to being logographic, and there is evidence that this quality was sometimes exploited in pre-modern multilingual texts.

³ The existence of language-independent elements has been noted for spoken communication in which they are referred to as *homophonous diamorphs*. Muysken (2000, 133) states that “when languages are similar or are perceived to be similar by bilingual speakers, switching is facilitated by specific words that sound alike in both languages” (Ter Horst and Stam 2018, 226). For example, prepositions *in* and *de* belong to several Indo-European languages. Could abbreviations have a similar function in written texts?

§24 The importance of abbreviations for multilingual practices is ignored by some digital approaches. King, Kübler and Hooper (2015) use automatic language identification on the *Chymistry of Isaac Newton*, an online resource which painstakingly encodes his alchemical symbols (Walsh and Hooper 2012). The approach used by King et al. is based on character n-grams. One of their aims is to identify language switch-points. Yet, they discard, among other things, the “large number of conventional alchemical symbols” used by Newton as “non-words” (536). While this is understandable from the point of view of the algorithm, as it operates on English written in the Latin alphabet, it does not take into account research into code-switching in which signs tending towards logography and semasiography can function as visual diamorphs and switch-points. Approaches to historical code-switching could be augmented by taking abbreviations into account.

4. Conditioned and unconditioned abbreviation

4.1. Conditioned variation

§25 An important point about abbreviation is that the abbreviated form is an alternative to the expanded form, which is used when a shorter form is needed under certain conditions. Many sources note this, albeit using different terminology and having slightly different theoretical emphases. Mazziotta (2008, §18) discusses how the two forms are independent signs linked by “synonymic relationship” which can be linked to “extralinguistic constraints.” Wittberger-Markwardt (2018, 68) and Nievergelt (2017, 231) call this relationship “lexical equivalence.” They further use the terms *Substituendum*, which refers to omitted full graphemes, and *Substitutum*, for ones that replace them (Wittberger-Markwardt 2018, 72). Samuels (1983); Thaisen and Da Rold (2009) and Thaisen (2011) use the terms “short form” and “long form,” including in short forms also shorter spellings (for example, *do* would be short form, but *doe* “do” or *doo* would be classified as long forms). They too note that shorter forms are used as an alternative for the longer ones. What all these authors agree upon is that the use of the shorter variant is subject to certain conditions.

§26 Variation that is conditioned by the surrounding environment could be seen as analogous to what in phonology is referred to as *conditioned variation*. I owe this observation to discussions with my former office mate Dr Raffaella Baechler (personal communication).

§27 Conditioned variation (2007) refers, for example, to sound changes, such as assimilation due to ease of articulation, in which a sound is affected by the phonological environment of the surrounding sounds. To use an example from the Concise Oxford Dictionary of Linguistics (2020), “the [m] of *comfort* can be described as a *conditioned variant*, in the position before a labiodental fricative, of the unit realized by another variant, [m], e.g. in *bumper*.” Thus, phonological variation is conditioned by the phonological environment of the surrounding sounds. Writing an abbreviated form instead of the full form is conditioned by the need to save time or space.

§28 Abbreviation is undertaken to save space or time, which Petti (1977, 22) calls *economy of space* and *economy of time*. Using an abbreviated form instead of the full one trades off some intelligibility (Avi-Yonah 1940, 9; Bozzolo et al. 1990; Cottureau 2005, 623), but it has the advantage of fitting the needed message into a smaller space, which could be important for reasons of conserving parchment or a carving surface of metal, stone or wood. Economy of time, on the other hand, was relevant throughout Antiquity and the Middle Ages, since “[t]he commonest way of committing words to writing was by dictating to a scribe” (Clanchy 1979, 97).⁴ In addition to these two, some authors, such as MacLean (2002), whose focus is on Greek epigraphy, mention saving labour. As he puts it, “abbreviations were used as a means of reducing labor and saving space on the stone’s surface” (2002, 49). Whatever the reason, the question then becomes whether quantitative palaeography can be used to study whether abbreviation was conditioned by something we can measure.

§29 Saving time may be somewhat difficult to establish hundreds of years later. It is impossible to measure how much time the scribe used to write a particular

⁴ Bozzolo et al. (1990, 18) make a further distinction between economy of time and space necessary for the transcription of a text (Economiser sur le temps et l’espace nécessaires à la transcription d’un texte.) and for “augmenting the speed of transcription to try and match it to that of speech” (augmenter la vitesse de transcription pour tenter de l’adapter à celle de la parole) and third playing on the length of the word to fit it into constraints of the line or of the page (Jouer sur la longueur matérielle des mots, en fonction des limites de l’espace scriptable: c’est-à-dire, dans le livre, “les contraintes de fin de ligne” et les “contraintes de fin de page”). However, it could be argued, that functions two and three broadly fit under economy of time and space respectively.

character, and there are many variables such as the type of script and its grade (Cottureau 2005, 623). Some scholars attempt this. Thaisen (2011) studies the use of shorter forms, including abbreviation, in copies of *Canterbury Tales*. He argues that they are more frequent in stretches of text which the scribe seems to have “completed late therefore tallies with the impression of the codicologists and palaeographers of a scribe finishing his work at some speed” (Thaisen 2011, 80). Cottureau (2005, 625) argues that conventionalised abbreviations, such as the *nomina sacra* as well as abbreviations of small function words could be associated with economy of time (Cottureau 2005, 628–629). There is thus arguably some evidence for economy of time in historical data. Nonetheless, it is difficult to measure and may be something that has to be argued for in the traditional qualitative way of making an argument.

§30 Economy of space, on the other hand, can be quantified and several studies have by now identified ways of measuring variation due to physical constraints. One of the easiest to quantify is simply the size of the manuscript in relation to abbreviation frequency. Not surprisingly, a number of studies that compare abbreviation frequencies to manuscript size (e.g. Cottureau 2005; Cottureau-Gabillet 2016; Thaisen 2011, 79; Honkapohja 2018, 250) have found abbreviation is more common in smaller manuscripts.

§31 Codicological units smaller than manuscript may also be constraints that contribute to abbreviation frequency. The need to conserve space may be particularly pressing towards the end of a quire, a page or a line. Camps (2016) tested for all of these variables but only found a fairly weak correlation for page and quire end. He found a more significant one for the end of the line. Shute (2017), who studied spelling variation in early printed books by William Caxton, using cluster analysis, noted that abbreviations are more likely to occur close to the right margin, and the results are statistically significant. The same phenomenon was also noted by Bozzolo et al. (1990, 23) as well as Cottureau (2005, 625–626), who also noted that the final word of the line is always more likely to be abbreviated than the first word of the line (627). She also found that the type of abbreviation may vary in different positions, as more specific abbreviations were more common towards the ends of lines as opposed to ones realised by simple nasals. There is thus ample quantitative evidence

for the importance of the end of the line as a conditioning variable for manuscript abbreviations.

§32 Going further down in specificity, we get to the context of individual words and letters. Things such as the length of the word and the number of syllables may have an effect of the use of abbreviation. Cottureau (2005, 653) found that final syllables were more likely to be abbreviated by more complex abbreviation signs, whereas the simple horizontal bar was more common in earlier syllables. One further discovery related to conditioning variables is the context of the preceding letter, made by Smith (2018, 205–206, 208, Figure 9.7), who found that the preceding letter form is the best predictor for word-final *f*: it occurs after *k*, *r*, *g*, *t* and *c* and not after *h*, *n*, *m*, *l*, *s* and *p*. Letter forms which terminate in a horizontal stroke are likely to be followed by the abbreviation, whereas ones terminated by lobes or vertical strokes are less likely followed by the abbreviated plural.

§33 Perhaps the most thorough consideration of economy of space comes from the pioneering French-Italian group *Quanticod*, who published a number of studies in quantitative diplomatics (see e.g. Bozzolo et al. 1990, 22). They approached abbreviation on the one hand with a detailed break-down based on what we here call conditioning variables, such as the position of the word in the line or the position of the syllable in the word, and on the other with a functional framework which differentiated between two types of economy of time (reading and writing) and economy of space. One of notions they proposed was *agrabilité* “agreeability” of certain syllables, which means that some types of syllables were more suitable for abbreviation than others. The work of *Quanticod* influenced later French scholarship, including Hasenohr (2002) as well as Cottureau-Gabillet (Cottureau 2005; Cottureau-Gabillet 2016) and Camps (2016), who have been able to demonstrate its suitability.

§34 There is a chance that agreeability was governed by something resembling prescriptive conventions. *Quanticod* was influenced by one fifteenth-century copy of *De laude scriptorium* by Jean Gerson, which contains an appendix called *Quedam regule de modo de titulandi*. The treatise contains what appear to be one fifteenth-century scribe’s “guidelines” or “stylesheet,” giving instructions, such as which abbreviations should be used in high grade books such as the Bible (Caen, Bibl. Mun.,

Coll. Mancel, ms. 131). It lists which abbreviations should be used for which syllables and where in the word they should be used.

§35 Further examples of abbreviation guidelines survive in early printed books. For example, German *Schryfftspiegel* (1527) also lists abbreviations by syllables. It instructs that abbreviation should never be used for majuscule and for minuscule only “in need” (“*in der not*”) at the end of the line (Römer 1997, 12–13). (Römer [13] also lists a number of other German early printed books which give instructions on the use of abbreviations.)

§36 To sum up, abbreviated forms are alternative spellings for fuller forms that were used under certain conditions: saving space, time and effort. Time and effort are difficult to study with quantitative precision, but saving space is one of the easier things to quantify. It is something where digital approaches have much to add to the argument. Abbreviation can then be analysed quantitatively by tagging and using as variables the context of the position in the line, the preceding and following, position in the word, the manuscript page and the quire, and performing a statistical analysis. If such causes can be pointed to as the reason why a scribe used an abbreviation, we are dealing with conditioned variation. If no clear, immediate reason for using an abbreviation rather than a full form can be found, we must ask the question of whether there are any other reasons which might have led to its use.

4.2. Unconditioned variation

§37 Economy of time and space do not, however, account for all mediaeval abbreviation. For example, Camps (2016, ccliv) notes that by carrying out a regression-using model which takes the ends of quires, pages or lines into account uncovers a number of weak correlations but leaves much variation unexplained. To continue with the analogy to phonology, such variation could be called unconditioned.

§38 Unconditioned or spontaneous variation in phonology refers to variation that cannot be attributed to the immediate phonetic context of the word. “Phonological variation is often studied from a sociolinguistic point of view, i.e. by examining the use of variants [...] such as sex, age, style, register and social class” (Anttila 2002, 206). These interact with the conditioned internal factors “such as

phonology, morphology, syntax and the lexicon,” which is “what makes it interesting to the phonologist.” Similarly, once we have sorted the internally conditioned variation caused by saving space, there could be much that abbreviations can tell us.

§39 Unconditioned variation for manuscript abbreviations could encompass variables related to geographic or diachronic variation or the socio-historical production circumstances of the manuscripts. Scribes from a certain area may have acquired their abbreviation practices from some local writing centre, such as a monastic scriptorium or an administrative office responsible for certain kinds of documents. It could also include diachronic change. Abbreviation practices change over time to adapt to different circumstances of text production, different writing materials, possibly even changes in reading practices, when literacy moved outside monasteries and spread to wider sections of the reading public.

§40 There have nevertheless been a number of quantitative surveys of abbreviations which have produced interesting results. Two of the earliest studies are by Bozzolo et al. (1990) and Hälvä-Nyberg (1988). Hälvä-Nyberg (1988) studied Latin inscriptions from Rome and Africa; extending from the earliest surviving sources to the eighth century, the corpus is selected to cover the switch from Roman pagan traditions to Christian ones. Bozzolo et al. (1990) studied fifteenth-century liturgical manuscripts. This was followed by Hasenohr (2002), who focused on French and Latin from the seventh to eleventh centuries. Römer (1997) contains a thorough quantitative description of a range of German manuscripts from southern Germany, Switzerland and Austria, starting from 1300. A recent project at the University of Zurich (see e.g. Wittberger-Markwardt 2018, Nievergelt 2017) focused on Old High German glosses. These should all provide some coverage of abbreviation practices in various areas and in various times, although they are not fully comprehensive.

§41 The question of how various socio-historical variables could influence the types of books copied was studied by Cottureau-Gabillet (Cottureau 2005, her PhD; Cottureau-Gabillet 2016, which contains some of the results in English). She studied hundreds of manuscripts from Paris libraries, which were selected with the criterion that the name of the copyist is known. Using sociohistorical variables in a statistical

enquiry, she found that abbreviation is less frequent in higher status copies. Higher abbreviation density can be predicted by the type of patron or higher “grade” of manuscript as more luxurious copies would have fewer.

§42 While French and German studies have been much more comprehensive with respect to abbreviation, the English and Scots historical dialectological tradition can make use of corpora based on manuscripts localised based on their language. These can be used to study abbreviations with respect to geographical variation. Such studies include Smith (2018), who found that the abbreviation *ʃ* “-is” is more common in more densely populated boroughs of Scotland: the scribes who copied more, were more likely to use it. Honkapohja (2019a and 2019b), on the other hand, discovered that there are a number of abbreviations specific to West Midlands counties in the Early Middle English period.

§43 There may also be differences within writing traditions that could be examined. Hasenohr (2002, 82–83) proposes that there is a major difference in monastic and scholastic abbreviation practices. Monastic writing was a slow and contemplative process. Scholastic abbreviation would include using more logographic abbreviation. Camps (2016, cclv) notes a number of interesting possibilities for this, proposing that abbreviations would peak with thirteenth-century scholasticism.

§44 Abbreviations can also give instructions on how to pronounce words in a number of ways, it has been argued. For example, Hasenohr (2002, 92) discusses superscript abbreviations in French in connection to pronunciation and whether the words were single or polysyllabic in Latin and French. N.R. Ker cites a manuscript in which the Latin word *neque* “neither,” which is expanded, has been consistently corrected to *neq*. He argues that the reason was stress in reading out aloud, as writing the word in full would mislead one to stressing the second syllable (cf. Clanchy 1979, 217; Ker 1960, 51).

5 Scribal profiles

§45 One of the most important uses of abbreviations in palaeography is for dating and localising scribal hands; for example, Ludwig Traube (1902) noted that when he is looking into the date of a manuscript, he immediately turns to the

abbreviations. To give a practical example, Ker (1960, 54) identifies one scribe as Norman, because he expands an Old English abbreviation (h) as “hoc” instead of “autem.” This leads to the question of whether abbreviations could be used digitally for the identification of scribal stints using methods developed for stylometry, compiling a scribal profile.

§46 There are pre-digital attempts to theorise and systematise the study of scribal accidentals. In the English tradition, this was proposed by McIntosh (1975), who hypothesises that every Middle English scribe is likely to have a unique linguistic profile (LP) of spellings and grammatical forms and also a graphetic profile (GP), of “linguistically sub-systemic [...] phenomena.” In French philology, Cazal, Parussa, Pignatelli and Trachsler (2003) have developed an equivalent system. A unique profile of graphetic elements is thus something that one would expect to be very well suited for computational analysis. There have been a number of interesting and encouraging results using digital approaches to scribal profiles, for example, by Thaisen (2020), who was able to use probabilistic modelling to identify scribal stints, de Bruijn and Kestemont (2013), who used n-grams and contrastive multivariate analysis to a Middle Low German chivalric romance.

§47 Out of the quantitative approaches to scribal profiles, there are few which make use of data which encodes abbreviations. Speed Kjeldsen (2013) notes differing abbreviation use by different scribes (391, 404). Kestemont (2015) studied scribal profiles, including abbreviations, using Principal Component Analysis for a richly encoded XML transcription of a letter collection of the Middle Dutch mystical female poet Hadewijch. He identified a change of scribal hand, based especially on the use of abbreviations, as “tilde abbreviations seem more common in the first part of the copy in ms. A,” while “B adopts a more abbreviation-rich orthography than A for this part of the letters.” (Kestemont 2015, 172–173). The result was thus that even though abbreviations were not the predominant focus of the study, they emerged as the distinguishing factor. This result, along with the discoveries based on graphetic variation, proves that there is definitely potential in the use of abbreviation as means for digital scribal profiling.

§48 A potential area of enquiry might be scribal profiles and high-frequency abbreviations. Kestemont (2015) noticed one scribe's propensity for using the very all-purpose tilde-abbreviation, which one scribe used much more than the other. Honkapohja (2019a: Figures 5–7) found that abbreviation in LAEME consists predominantly of certain types, especially the macron and hook. Camps (2016, cclvi) proposed that the best approach for studying similarities and differences between manuscript witnesses of the same text would be to focus on certain abbreviation types. A quantitative profiling of these high-function types is something which definitely deserves more investigation.

§49 An alternative would be to focus on function word and content words. Abbreviations seem to have been particularly frequent for function words. For example, Hasenohr (2002, 80) notes that most of these abbreviations were created in the twelfth and thirteenth centuries for use in cursive writing; they mainly affect the endings, the adverbs, particles and pronouns, as well as the forms of the verb *esse* –words “which come often under the pen” (*qui reviennent souvent sous la plume*). Rogos-Hebda (2018, 55) notes differences in the use of abbreviations by two Chaucer scribes. Honkapohja (2018, 267–8) notes that scribes are more likely to copy lexical words directly, but have individual profiles for function words. A focus on function words has parallels with developments in stylometry in which function words were considered to be the best way of establishing authorship profiles, before n-grams proved to be more efficient (Kestemont 2014, 62). Nevertheless, with abbreviations the use of function words has not been compared to n-grams, which are also partly popular as they can be used with plain text (De Bruijn and Kestemont 2013, 182). Perhaps a rich transcription in which abbreviations are encoded could complement the study of scribal stints by n-grams. Or perhaps a transcription in which abbreviations are encoded could itself be subjected to n-gram based analysis.

6 Encoding abbreviations

§50 In order to study abbreviations quantitatively, we need to count them somehow. While earlier studies collected their dataset manually, these days the

standard is TEI P5 XML. TEI P5 is a flexible framework divided into modules, which provide a number of alternative ways to encode abbreviations. As we have now had a few decades' worth of projects using TEI XML, the problems, solutions and theoretical implications of dealing with different types of abbreviations have been discussed in many sources for different languages and document types (cf. Heiden et al. 2002; Driscoll 2006, 2009; Mazziotta 2008; Stutzmann 2010, 2014, 284; Speed Kjeldsen 2013, 34–38; Honkapohja 2013a; Horst and Stam 2018). Some of them decide to encode abbreviations in their resources and, as a by-product, publish an article or guidelines detailing their editorial choices. Moreover, individual projects and even individuals working within the same project may encode the same abbreviation differently. The aim of this paper is not to argue in favour of any particular encoding choices, but to function as a review article that will help readers to navigate these waters and also to outline how the theoretical contribution made in the previous section fits with two prominent encoding systems. However, one particular problem deserves highlighting as it is connected with the writing system cline discussed in section 3, as it offers a greater theoretical clarity on it.

§51 The theoretical distinction outlined in section three fits in well with the discussion by Driscoll (2006, 2009), which is influential not only in Scandinavian but also Anglophone scholarship. Some abbreviations replace a string of characters, while others are more logographic. The type of the abbreviation and whether it has a clear referent in a string of Latin alphabetical characters will affect the “editorial policy” on how it is best to be encoded. It is necessary to distinguish between abbreviations that refer to the entire word and ones that correspond to a sequence of characters. Driscoll (2006, 259) calls these abbreviations “with a graphemic reference (superscript letters and signs and the remainder of the brevigraphs)” and those “with a lexical reference (suspensions, contractions, and a number of brevigraphs).” He goes on to say that “[i]t strikes one as counterintuitive to treat the former on anything other than the whole-word level, while treating the latter in the same way seems equally misconceived.” This distinction corresponds exactly with (attempted) phonographic and logographic writing systems (see Examples 1 and 2).


```

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

```

Example 1: Abbreviation with graphemic/phonographic reference.

```

<w>
  <choice>
    <abbr>
      <am>&amp;</am>
    </abbr>
    <expn>
      <ex>and</ex>
    </expn>
  </choice>
</w>

```

Example 2: Abbreviation with lexical/logographic reference.

§52 A different solution is proposed by Mazziotta (2008) and discussed by Stutzmann (2010 and 2013). This approach advocates a different way of seeing abbreviations from the point of view of the writing system, focusing on the work of the scribe. The system seeks to model the scribe's strokes in a "descriptive" way based on their position in the "graphic space" without reference to the way they are vocalised in a phonological word (Mazziotta 2008, §13, §18). Mazziotta divides them into a few axes (§19) and proposes terminology for describing various relationships (cf. Mazziotta 2008, §18–§49). For example, the very common "crossed-p" abbreviation *p* is not seen as a single sign, but rather the letter *p* modified by a *cénégram* (Stutzmann 2008, 265–66). The solution he proposes is elegant, but somewhat complex (it

results in quite a bit of terminology, something which writing systems research is notorious for (Powell 2009, 263); Mazziotta uses the terms *périgramme*, *linégramme*, *cénégramme*, *caténogramme*, *logogram* and *plérégramme*) and, as Stutzmann (ibid.) notes, goes against the encoding philosophy of MUFI and Unicode. It does have advantages for encoding abbreviations, as it removes the common transcription problem of trying to fit into exact encoding boxes something which was a single stroke for the scribe. That is, in ambiguous cases, whether a horizontal bar which crosses through, e.g., a tall ascender for l but also simply sits on top of other letters, is a “macron” or a “crossed-l.” Encoding based on these theoretical conceptions is discussed by Mazziotta (2008, 3.2.1–3.2.2) and Stutzmann (2010 and 2013). In the form applied by the ORIFLAMMS (2020) project it would be as follows. The encoding would be like the one presented in Example 3.

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

Example 3: ORIFLAMMS mark-up for abbreviations.

§53 The views presented in this article are not incompatible with the approach advocated by Mazziotta and Stutzmann. The various types of combining marks fall under what Powell (2008, 45–46) calls “auxiliary signs and devices,” placing them in the same group with features such as capitalisation, diacritic marks or punctuation, which are established conventions developed on top of the writing system. On the other hand, the discussion is on a different level. This paper deals with the graphemic level and focuses on the reader. Mazziotta’s system models them from the point of view of scribal practice. He even makes it clear that his concern is not with the

psychology of reading (2008, §46). ORIFLAMMS (2020) treats them in addition to various graphetic features. (Moreover, as Mazziotta admits his theoretical approach solves abbreviations from a theoretical point of view of strokes made by the scribe described in relation to other strokes. However, it is not fully compatible with the psychology of reading – as abbreviated forms become kind of logograms. Experienced readers are likely to read on word-level rather than individual graph level. Something which Hasenohr suggested as a transformation from the contemplative reading of monasteries to the logographic reading in universities). While there are advantages to graphetic transcription, such as being able to carry out statistical investigations without presupposing a writing system, the question of referent for abbreviations is more complex than for some other features and is of linguistic interest also from the point of view of where these features fit in the typology of writing systems from a graphemic point of view.

§54 One further problem related to encoding abbreviations derives from the language-independence of especially some logograms (see section 3.2 above). Ter Horst and Stam also discuss encoding visual diamorphs in XML. Among other things, the `@lang` can be used to specify which language a certain abbreviated word belongs to. The mark-up for visual diamorphs is addressed by Ter Horst and Stam (2018), who add a tag for words that are part of both languages. “The preferred method for signalling code-switches in XML is the language-attribute (`@lang=""`). Apart from the standard value for Latin ("`la`") and Irish ("`ga`", " for *Gaelic*) we added the custom value '`ga-la`' for visual dimorphs" (Ter Horst and Stam 2018, 224). TEI Guidelines are thus well-suited for handling not only the logography/phonography variations, but also visual diamorphs.

7 Suitable resources for the study of abbreviations

§55 Studying abbreviations quantitatively requires digital transcriptions that encode the forms of abbreviation as well as their expansions. While many earlier corpora were based on printed books, the situation is improving. The numbers are not nearly as high as corpora that do not encode abbreviations or manuscript images online (Robinson 2016, 182–7), but there are now several resources available that

offer diplomatic transcriptions that indicate abbreviations in a format suitable for quantitative analysis.

§56 Editions of individual works include Murchinson (2017), Camps (2016), Dunning (2016), Honkapohja (2013b) and De Leeuw van Weenen (2009). All of these make use of the TEI architecture for abbreviations, and allow toggling between diplomatic view and normalised view, which means the abbreviations are available to the user. Larger corpora that encode abbreviations in a way that enables applying quantitative methods include texts made available by the *Medieval Nordic Text Archive* (MENOTA), a network of libraries, archives and research departments of Old Icelandic, Old Norwegian and Old Swedish Texts. The project makes use of TEI P5 XML. Another one is the Base de Français Médiéval (BFM 2020) which contains French texts from the ninth to the fifteenth centuries tagged in TEI P5 XML. *The Chymistry of Isaac Newton* (2020) makes several of Newton's alchemical writings available, allowing access both to diplomatic and normalised versions. *Early English Books Online* (EEBO 2020) does encode abbreviations, but not their expansions, as does the Helsinki Corpus TEI XML version. Abbreviations in this approach are very much seen as part of the allographic system and the aim of the project ORIFLAMMS (2020), which applies it, was to come up with a comprehensive typology for medieval writing systems. Several teams contributed to ORIFLAMMS. The project has published large corpora and made lists of abbreviations available on GitHub.

§57 There are also resources that do not use TEI encoding. The digital edition *An Electronic Text Edition of Depositions 1560–1760* (Kytö, Grund, and Walker 2011) uses XML with a high degree of diplomatic accuracy but is not TEI-based. In English studies, there are a number of corpora that encode abbreviations in expanded form, using an ASCII-based system originating in the 1990s. There include the Edinburgh resources *Linguistic Atlas of Early Middle English* (LAEME) and *Linguistic Atlas of Older Scots* (LAOS), as well as the *Middle English Grammar Corpus* (MEG-C) and *Middle English Local Documents* (MELD) corpora compiled at the University of Stavanger.

§58 A more recent addition are developments in deep-learning algorithms and *Handwritten Text Recognition* (HTR) used for projects like HIMANIS (2020)

and *Transkribus*. These tools are able to provide plain text querying and indexing of thousands of pages of medieval manuscripts using image data and enabling the use of distant-reading techniques to manuscript data. As they work on images, the algorithms are able to retrieve both abbreviated and unabbreviated strings in different volumes and different handwritings. Thus, developments in HTR and deep-learning promise to revolutionise enabling the quantitative study of abbreviations with “big data” in the not-too-distant future. Yet as Stutzmann et al. (2018) note, at the moment, there is still some work to be done before “uneven, automatically generated data” can be used as a reliable research tool. The risk is that automatic processes might generate huge quantities of bad data. Nevertheless, for scholars wanting to study abbreviations quantitatively there are now several resources available and the situation is likely to improve further. With the increasing availability of resources that enable quantitative palaeography, we might look towards something of a golden age for this fascinating area of study.

Acknowledgements

This research has been funded by the Swiss National Science Foundation (SNF). I am grateful for their support. I am also grateful for Sara Norja for her language check and the anonymous reviewers for their comments.

Competing interests

The author has no competing interests to declare.

Contributors

Editorial

Recommending editor

Lisa Fagin Davis, Medieval Academy of America, U.S.A.

Recommending referees

Jenna Mead, University of Western Australia, Australia

Peter Robinson, University of Saskatchewan, Canada

Section, copy, and layout editor

Shahina Parvin, The Journal Incubator, University of Lethbridge, Canada

References

- Andrews, Tara,** and **Caroline Macé.** 2013. "Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation through Graph Analysis of Texts and Stemmata." *Literary and Linguistic Computing* 28(4): 504–21. DOI: <https://doi.org/10.1093/lc/fqt032>
- Anis, Jacques.** 2007. "Neography: Unconventional Spelling in French SMS Text Messages." In *The Multilingual Internet: Language, Culture, and Communication Online*, edited by Brenda Danet and Susan C. Herring, 87–115. Oxford Scholarship Online. DOI: <https://doi.org/10.1093/acprof:oso/9780195304794.003.0004>
- Anttila, Arto.** 2002. "Variation and Phonological Theory." In *Handbook of Language Variation and Change*, edited by Jack Chambers, Peter Trudgill and Natalie Schilling-Estes, 206–243. Oxford: Blackwell. DOI: <https://doi.org/10.1111/b.9781405116923.2003.00013.x>
- Avi-Yonah, Michael.** 1940. *Abbreviations in Greek Inscriptions (The Near East, 200 B.C.—A.D. 1100)*. London: Milford.
- Benskin, Michael.** 1977. "Local Archives and Middle English Dialects." *Journal of the Society of Archivists* 5: 500–14. DOI: <https://doi.org/10.1080/00379817709514090>
- BFM.** 2020. Base de français médiéval. Edited by Céline Guillot-Barbance, Serge Heiden, Alexei Lavrentiev, Matthieu Decorde, and Nadine Pontal. Accessed September 30. <http://bfm.ens-lyon.fr/spip.php?article339>.
- Bieswanger, Markus.** 2007. "2 Abbrevi8 or Not 2abbrevi8: A Contrastive Analysis of Different Space- and Time-saving Strategies in English and German Text Messages." In *Texas Linguistic Forum 50*, edited by Simeon Floyd, Taryne Hallet, Sae Oshima, and Aaron Shield. <https://didattica.uniroma2.it/assets/uploads/corsi/39540/smsgermanenglish.pdf>.
- Bozzolo, Carla, Dominique Coq, Denis Muzerelle,** and **Ezio Ornato.** 1990. "Les Abréviations dans les Livres Liturgiques du XVe Siècle: Pratique et Théorie." In *Actas del VIII coloquio del Comité internatiocional de paleografía Latina*, Madrid-Toledo, Sept.–Oct. 1987, 17–27. Madrid: Joyas Bibliográficas.

- Camps, Jean-Baptiste.** 2016. "La `Chanson d'Otinel': édition complète du corpus manuscrit et prologomènes à l'édition critique." PhD Thesis, Paris: Paris-Sorbonne. DOI: <https://doi.org/10.4000/peme.13004>
- Cappelli, Adriano.** 1990 [1899]. *Lexicon Abbreviaturarum Dizionario Di Abbreviature Latine Ed Italiane*. Milano: Hoepli.
- Cazal, Yvonne, Gabriella Parussa, Cinzia Pignatelli, and Richard Trachsler.** 2003. "L'orthographe: du manuscrit médiéval à la linguistique modern." *Médiévales* 45: 99–118. DOI: <https://doi.org/10.4000/medievales.969>
- Chymistry of Isaac Newton, The.** 2020. Accessed September 30. <http://webapp1.dlib.indiana.edu/newton/>.
- Clanchy, Michael T.** 1979. *From Memory to Written Record: England 1066–1307*. London: Edward Arnold.
- Conditioned Variation.** 2007. *The Concise Oxford Dictionary of Linguistics*. 2nd ed. Oxford: OUP. Accessed September 30, 2020. <https://www.oxfordreference.com/view/10.1093/acref/9780199202720.001.0001/acref-9780199202720-e-633>.
- Cook, Vivian, and Des Ryan.** 2016. "Background to the English Writing System." In *The Routledge Handbook of the English Writing System*, edited by Vivian Cook and Des Ryan, 5–24. London and New York: Routledge. DOI: <https://doi.org/10.4324/9781315670003>
- . 2016. *The Routledge Handbook of the English Writing System*. London and New York: Routledge. DOI: <https://doi.org/10.4324/9781315670003>
- Cottureau, Emilie.** 2005. "La copie et les copistes français de manuscrits aux XIVE et XVe siècles. Etude sociologique et codicologique." PhD Thesis, Paris: Paris-Sorbonne.
- Cottureau-Gabillet, Emilie.** 2016. "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), edited by Sébastien Barret, Dominique Stutzmann, and Georg Vogeler, 129–63. Utrecht Studies in Medieval Literacy 35. Turnhout: Brepols. DOI: <https://doi.org/10.1484/M.USML-EB.5.112434>

- Czajkowski, Luise.** 2018. "Zur Interpretation von Kürzungsstrichen in Handschriften des 14. und 15. Jahrhunderts." In *Handschriften als Quellen der Sprach- und Kulturwissenschaft: Aktuelle Fragestellungen – Methoden – Probleme*, edited by Anette Kremer, Vincenz Schwab, 83–98. Bamberg: University of Bamberg Press.
- Daniels, Peter T., and William Bright.** 1996. *The World's Writing Systems*. New York/Oxford: OUP.
- De Bruijn, Elisabeth, and Mike Kestemont.** 2013. "Contrastive Multivariate Analyses of the Middle Low German Flos unde Blankeflos." *Neuphilologische Mitteilungen* 114(2): 171–205.
- De la Cruz-Cabanillas, Isabel, and Irene Diego-Rodriquez.** 2018. "Abbreviations in Medieval Medical Manuscripts." *SELIM: Journal of the Spanish Society for Medieval English Language and Literature. Revista de la Sociedad Española de Lengua y Literatura Inglesa Medieval* 23: 163–183.
- De Leeuw van Weenen, Andrea.** 2009. *Alexanders Saga: AM 519A 4 in the Arnarnagnæan Collection, Copenhagen*. Copenhagen: Museum Tusulanum Press.
- Driscoll, Matthew James.** 2006. "Levels of Transcription." In *Electronic Textual Editing*, edited by Lou Burnard, Katherine O'Brien O'Keefe and John Unsworth, 254–261. New York: Modern Language Association.
- . 2009. "Marking up Abbreviations in Old Norse-Icelandic Manuscripts." In *Medieval Texts – Contemporary Media: The Art and Science of Editing in the Digital Age*, edited by Maria Grazia Saibene and Marina Buzzoni, 13–34. Pavia: Ibis.
- . 2010. "The Words on the Page: Thoughts on Philology, Old and New." In *Creating the Medieval Saga: Versions, Variability, and Editorial Interpretations of Old Norse Saga Literature*, edited by Judy Quinn and Emily Lethbridge, 87–104. Odense: University of South Denmark.
- Dunning, Andrew.** 2016. "Hugh of Saint Victor's De Quinque Septenis (On the Five Sevens) and Its Versification in Samuel Presbiter's De Oratione Dominica (On the Lord's Prayer)." *Scholarly Editing: The Annual of the Association of Documentary Editing* 37. Accessed September 30, 2020. <http://scholarlyediting.org/2016/editions/intro.dunning.html>.

- EEBO.** 2020. Early English Books Online. Accessed September 30. <https://quod.lib.umich.edu/e/eebogroup/>.
- eLALME.** 2013 (1986). *An Electronic Version of a Linguistic Atlas of Late Mediaeval English*. Edited by Michael Benskin, Margaret Laing, Vasilios Karaiskos and Keith Williamson. Aberdeen: Aberdeen University Press. Accessed September 30, 2020. <http://www.lel.ed.ac.uk/ihd/elalme/elalme.html>.
- Gelb, Ignace J.** 1952. *A Study of Writing*. Chicago: University of Chicago Press.
- GitHub.** 2020. "Oriflamms." Accessed September 30. <https://github.com/oriflamms>.
- Greg, W. W.** 1951. "The Rationale of Copy-Text." *Studies in Bibliography* 3: 19–36.
- Hälvä-Nyberg, Ulla.** 1988. Die Kontraktionen auf den Lateinischen Inschriften Roms und Afrikas: bis zum 8. Jh. N.Chr. Helsinki: Suomalainen tiedeakatemia.
- Hasenohr, Geneviève.** 2002. "Écrire en latin, écrire en roman: réflexions sur la pratique des abréviations dans les manuscrits français des XIIe et XIIIe siècles." In *Langages et peuples d'Europe: cristallisation des identités romanes et germaniques (VIIe-XIe siècle)*, edited by Michel Banniard, 79–110. Toulouse: CNRS Université de Toulouse-Le Mirail.
- Hector, L. C.** 1958. *The Handwriting of English Documents*. Ilkley: Scolar Press.
- Heiden, Serge, Céline Guillot, Alexei Lavrentiev, and Lauranne Bertrand.** 2002. "Manuel d'encodage XML-TEI Des Textes de La Base de Français Médiéval." Accessed September 30, 2020. http://bfm.ens-lyon.fr/IMG/pdf/Manuel_Encodage_TEI.pdf.
- Heimann, David, and Kay, Richard.** 1982. *The Elements of Abbreviation in Medieval Latin Paleography* (University of Kansas Publications 52). Kansas: University of Kansas Libraries. Accessed December 02, 2020. <https://kuscholarworks.ku.edu/bitstream/handle/1808/1821/47cappelli.pdf>.
- Helsinki Corpus TEI XML Edition.** 2011. First edition. Designed by Alpo Honkapohja, Samuli Kaislaniemi, Henri Kauhanen, Matti Kilpiö, Ville Marttila, Terttu Nevalainen, Arja Nurmi, Matti Rissanen and Jukka Tyrkkö. Implemented by **Henri Kauhanen** and **Ville Marttila**. Based on The Helsinki Corpus of English Texts (1991). Helsinki: The Research Unit for Variation, Contacts and

Change in English (VARIENG), University of Helsinki. Accessed December 01, 2020. <https://metashare.csc.fi/repository/browse/helsinki-corpus-tei-xml-edition-2011-korp/12e568a88d3511e7a649005056be118e4bbf799466094fd a8c7b8bdcc7e1020f/>.

HIMANIS. 2020. The HIMANIS Project: Indexing the Trésor des Chartres Registers. Accessed September 30. <https://himanis.org/>.

Honkapohja, Alpo. 2013a. "Manuscript Abbreviations in Latin and English: History, Typologies and How to Tackle Them in Encoding." In *Studies in Variation, Contacts and Change in English Volume 14: Principles and Practices for the Digital Editing and Annotation of Diachronic Data*, edited by Anneli Meurman-Solin and Jukka Tyrkkö. Accessed September 30, 2020. <http://www.helsinki.fi/varieng/series/volumes/14/honkapohja/>.

---. 2013b. "The Trinity Seven Planets." *Scholarly Editing: The Annual of the Association of Documentary Editing* 34. Accessed September 30, 2020. <http://scholarlyediting.org/2013/editions/intro.sevenplanets.html>.

---. 2017. *Alchemy, Medicine, and Commercial Book Production: A Codicological and Linguistic Study of the Voigts-Sloane Manuscript Group (Texts and Transitions, 9)*. Turnhout: Brepols. DOI: <https://doi.org/10.1484/M.TT-EB.5.111529>

---. 2018. "'Latin in Recipes?' A Corpus Approach to Scribal Abbreviations in 15th-Century Medical Manuscripts." In *Multilingual Practices in Language History: English and beyond*, edited by Päivi Pahta, Janne Skaffari and Laura Wright, 243–271. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9781501504945-012>

---. 2019a. "Anchorites and Abbreviations: A Corpus Study of Abbreviations of Germanic and Romance lexicon in the Ancrene Wisse." In *Current Explorations in Middle English: Selected Papers from the 10th International Conference on Middle English (ICOME), University of Stavanger (Studies in English Medieval Language and Literature, Vol 56), Norway, 2017*, edited by Merja Stenroos, Martti Mäkinen, Kjetil Vikhamar Thengs and Oliver Martin Traxel, 35–64. Bern: Peter Lang.

---. 2019b. "The Brief Rise and Fall of a Superscript Abbreviation for the First Person Singular Pronoun in the West Midlands Between 1250 and 1500." *Notes & Queries* 66(4): 502–508. ISSN: 0029-3970. DOI: <https://doi.org/10.1093/notesj/gjz112>

- Honkapohja, Alpo, and Aino Liira.** 2020. "Abbreviations and Standardisation in the Polychronicon: Latin to English, and Manuscript to Print." In *The Multilingual Origins of Standard English (MOSTE)*, edited by Laura Wright, 269–316. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110687545-010>
- Honkapohja, Alpo, Samuli Kaislaniemi, and Ville Marttila.** 2009. "Digital Editions for Corpus Linguistics: Representing Manuscript Reality in Electronic Corpora." In *Corpora: Pragmatics and Discourse. Papers from the 29th International Conference on English Language Research on Computerized Corpora (ICAME 29)*. Ascona, Switzerland, May 14–18, 2008, edited by Andreas Jucker, Daniel Schreier and Marianne Hundt, 451–474. Amsterdam: Rodopi.
- Ker, Neil R.** 1960. *English Manuscripts in the Century after the Norman Conquest*. Oxford: Claredon.
- Kestemont, Mike.** 2014. "Function Words in Authorship Attribution. From Black Magic to Theory?" In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, edited by Anna Feldman, Anna Kazantseva and Stan Stepakowicz, 59–66. Stroudsburg: Association for Computational Humanities. Accessed September 30, 2020. <https://www.aclweb.org/anthology/volumes/W14-09/>. DOI: <https://doi.org/10.3115/v1/W14-0908>
- . 2015. "A Computational Analysis of the Scribal Profiles in Two of the Oldest Manuscripts of Hadewijch's Letters." *Scriptorium* 69: 159–75.
- King, Levi, Sandra Kübler, and Wallace Hooper.** 2015. "Word-level Language Identification in the Chymistry of Isaac Newton." *Literary and Linguistic Computing* 30(4): 532–540. DOI: <https://doi.org/10.1093/lc/fqu032>
- Kopaczyk, Joanna.** 2011. "A V or not a V? Transcribing Abbreviations in Seventeen Manuscripts of the 'Man of Law's Tale' for a Digital Edition." In *Scribes, Printers, and the Accidentals of their Texts*, edited by Jacob Thaisen and Hanna Rutkowska, 91–106. Frankfurt: Peter Lang.
- Kwakkel, Erik.** 2012. "Biting, Kissing and the Treatment of Feet: The Transitional Script of the Long Twelfth Century." In *Turning Over a New Leaf: Change and Development in the Medieval Book*, edited by Erik Kwakkel, Rosamund McKitterick and Rodney Thomson, 79–124. Leiden: Leiden University Press.

- Kytö, Merja, Peter Grund, and Terry Walker.** 2011. *Testifying to Language and Life in Early Modern England: Including CD-ROM: An Electronic Text Edition of Depositions 1560–1760 (ETED)*. Amsterdam: Benjamins. DOI: <https://doi.org/10.1075/z.162>
- LAEME.** 2013. A Linguistic Atlas of Early Middle English, 1150–1325, Version 3.2. Edited by Margaret Laing, Michael Benskin, Margaret Laing, Vasilios Karaïskos, and Keith Williamson. Edinburgh: The University of Edinburgh. Accessed September 30, 2020. <http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html>.
- LAOS.** 2020. Linguistic Atlas of Older Scots, Phase One: 1380–1500. Edited by Keith Williamson. Edinburgh: The University of Edinburgh. Accessed September 30. <http://www.lel.ed.ac.uk/ihd/laos1/laos1.html>.
- Lass, Roger.** 2004. “Ut Custodiant Litteras: Editions, Corpora and Witnesshood.” In *Methods and Data in English Historical Dialectology*, edited by Marina Dossena and Roger Lass, 21–48. Bern: Peter Lang.
- Lowe, Kathryn.** 2006. “Palaeography.” *Encyclopedia of Language and Linguistics*. 2nd ed., edited by Keith Brown, 134–141. London: Elsevier. DOI: <https://doi.org/10.1016/B0-08-044854-2/04876-8>
- MacLean, Bradley H.** 2002. *An Introduction to Greek Epigraphy of the Hellenistic and Roman Periods from Alexander the Great Down to the Reign of Constantine (323 B.C.-A.D. 337)*. Ann Arbor: University of Michigan Press.
- Mazziotta, Nicolas.** 2008. “Traiter Les Abréviations Du Français Médiéval. Théorie de l’écriture et Pratiques d’encodage.” *Corpus 7*. <http://corpus.revues.org/index1517.html>. DOI: <https://doi.org/10.4000/corpus.1517>
- McIntosh, Angus.** 1975. “Scribal Profiles from Middle English texts.” *Neuphilologische Mitteilungen* 76: 218–35.
- MEG-C.** 2011. *The Middle English Grammar Corpus. Version 2011.1*. Compiled by Merja Stenroos, Martti Mäkinen, Simon Horobin and Jeremy Smith, University of Stavanger. Accessed September 30, 2020. <https://www.uis.no/research/history-languages-and-literature/the-mest-programme/the-middle-english-grammar-corpus-meg-c/how-to-cite-us/>.

- MELD.** 2017. *A Corpus of Middle English Local Documents. Version 2017.1.* Compiled by Merja Stenroos, Kjetil Thengs and Geir Bergström, University of Stavanger. Accessed December 02, 2020. <http://www.uis.no/research/history-languages-and-literature/the-mest-programme/a-corpus-of-middle-english-local-documents-meld/>.
- Menota.** 2020. "Medieval Nordic Text Archive." Accessed September 30. <https://www.menota.org/forside.xhtml>.
- Murchinson, Krista A.** 2017. "'The Effects of the Seven Sins': A Critical Edition." *Scholarly Editing: The Annual of the Association of Documentary Editing* 38. Accessed September 30, 2020. <http://scholarlyediting.org/2017/editions/sevensins/intro.html>.
- Muysken, Pieter.** 2000. *Bilingual Speech: A Typology of Code-mixing.* Cambridge: Cambridge University Press.
- Nievergelt, Andreas.** 2017. "Glossen aus einem einzigen Buchstaben." In *The Annotated Book in the Early Middle Ages: Practices of Reading and Writing*, edited by Mariken Teeuwen and Irene van Renswoude, 285–304. Turnhout: Brepols. DOI: <https://doi.org/10.1484/M.USML-EB.5.115024>
- Page, Raymond Ian.** 1960. *Gibbons saga, Editiones Arnemagnaeanae B.2.* Copenhagen: Munksgaard.
- Petti, Anthony G.** 1977. *English Literary Hands from Chaucer to Dryden.* London: Arnold.
- Powell, Barry B.** 2009. *Writing: Theory and History of the Technology of Civilization.* Chichester: Wiley-Blackwell.
- Roberts, Jane.** 2005. *Guide to Scripts Used in English Writing up to 1500.* London: British Library.
- Robinson, Peter.** 2016. "The Digital Revolution in Scholarly Editing." In *Ars Edendi Lecture Series, Vol. IV*, edited by Barbara Crostini, Gunilla Iversen and Brian M. Jensen, 181–207. Stockholm: Stockholm University Press. DOI: <https://doi.org/10.16993/baj.h>
- Robinson, Peter, and Elizabeth Solopova.** 1993. "Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue." In *The Canterbury Project*

Occasional Papers volume I, edited by Norman F. Blake and Peter Robinson, 19–52. Oxford: Office for Humanities Communication.

- Rogos, Justyna.** 2011. "On the Pitfalls of Interpretation: Latin Abbreviations in MSS of the Man of Law's Tale." In *Foreign Influences on Medieval English*, edited by Jacek Fisiak and Magdalena Bator, 47–54. Bern: Peter Lang.
- . 2012. "Isles of Systemacity in the Sea of Prodigality? Non-alphabetic Elements in Manuscripts of Chaucer's 'Man of Law's Tale.'" Accessed September 30, 2020. <https://docplayer.net/41316899-Isles-of-systematicity-in-the-sea-of-prodigality-non-alphabetic-elements-in-manuscripts-of-chaucer-s-man-of-law-s-tale-justyna-rogos.html>.
- Rogos-Hebda, Justyna.** 2018. "Text and Image: Revisiting Traube's *Halbgraphische Objekte* in a Paleographic-Pragmatic Approach to Scribal Abbreviation" *Kwartalnik Neofilologiczny*, LXV, 1/2018.
- Römer, Jürgen.** 1993. "Les Recherches sur les Abréviations dans les Textes en Langue Vulgaire, Notamment Germaniques." *Gazette du livre médiéval* 22(1993): 7–13. Accessed September 30, 2020. DOI: <https://doi.org/10.3406/galim.1993.1226>
- Römer, Jürgen.** 1997. *Geschichte der Kürzungen: Abbréviaturen in deutschsprachigen Texten des Mittelalters und der Frühen Neuzeit*. Göttingen: Kümmerle Verlag.
- Samuels, Michael.** 1983. "The Scribe of the Hengwrt and Ellesmere Manuscripts of the Canterbury Tales." *Studies in the Age of Chaucer* 5: 49–65. DOI: <https://doi.org/10.1353/sac.1983.0002>
- Shute, Rosie.** 2017. "Pressed for Space: The Effects of Justification and the Printing Process on Fifteenth-Century Orthography." *English Studies* 98(3): 262–82. DOI: <https://doi.org/10.1080/0013838X.2017.1250197>
- Smith, Daisy.** 2018. "The Predictability of {-S} Abbreviation in Older Scots Manuscripts According to Stem-final Littera." In *Historical Dialectology in the Digital Age*, edited by Rhona Alcorn, Joanna Kopaczyk, Bettelou Los, and Benjamin Molineaux, 187–211. Edinburgh: Edinburgh University Press.
- Speed Kjeldsen, Alex.** 2013. *Filologiske Studier I Kongesagahåndskriftet Morkinskinna*. Copenhagen: Museum Tusulanum Press.

- Stokes, Peter.** 2014. *English Vernacular Minuscule from Æthelred to Cnut, circa 990 – Circa 1035*, Publications of the Manchester Centre for Anglo-Saxon Studies 14. Woodbridge: Boydell and Brewer.
- Stutzmann, Dominique.** 2010. "Paléographie Statistique pour Décrire, Identifier, Dater... Normaliser pour Coopérer et Aller plus Loin?" In *Kodikologie und Paläographie im digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2*, Schriften des Instituts für Dokumentologie und Editorik 3., edited by Franz Fischer, Christiane Fritze and Georg Vogeler, 247–77. Norderstedt: BoD.
- . 2013. "Ontologie des Formes et Encodage des Textes Manuscrits Médiévaux. Le Projet ORIFLAMMS." *Document numérique* 16(3): 81–95. DOI: <https://doi.org/10.3166/dn.16.3.69-79>
- . 2014. "Conjuguer Diplomatie, Paléographie et Édition Électronique : les Mutations du XIIe Siècle et la Datation des Écritures par le Profil Scribal Collectif." In *Digital Diplomats. The Computer as a Tool for the Diplomatist?*, edited by Antonela Ambrosio, Sébastien Barret and Georg Vogeler. Archiv für Diplomatik. Beiheft 14, 271–90. Vienna, Cologne, Weimar: Böhlau Verlag. DOI: <https://doi.org/10.7788/boehlau.9783412217020.271>
- Stutzmann, Dominique, Christopher Kermorvant, Enrique Vidal, Sukalpa Chanda, Sébastien Hamel, Joan Puigcerver Pérez, Lambert Schomaker, and Alejandro H. Toselli.** 2018. "Handwritten Text Recognition, Keyword Indexing, and Plain Text Search in Medieval Manuscripts." Conference paper presented at *Digital Humanities 2018 Conference*, Mexico City, June 26–29. Accessed September 30, 2020. <https://dh2018.adho.org/handwritten-text-recognition-keyword-indexing-and-plain-text-search-in-medieval-manuscripts>.
- Ter Horst, Tom, and Nike Stam.** 2018. "Visual Diamorphs: The Importance of Language Neutrality in Code-Switching from Medieval Ireland." In *Multilingual Practices in Language History: English and Beyond*, edited by Päivi Pahta, Janne Skaffari and Laura Wright, 199–222. Berlin: De Gruyter.
- Thaisen, Jacob.** 2011. "Adam Pinkhurst's Short and Long Forms." In *Scribes, Printers, and the Accidentals of their Texts*, edited by Jacob Thaisen and Hanna Rutkowska, 73–90. Frankfurt: Peter Lang.

- . 2017. "Secretary Letter Forms in Country Durham." *Folia Linguistica*, 38: 263–280. DOI: <https://doi.org/10.1515/flih-2017-0009>
- . 2020. "Standardisation, Exemplars, and the Auchinleck manuscript". *The Multilingual Origins of Standard English*, edited by Laura Wright, 165–190. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9783110687545>
- Thaisen, Jacob, and Orietta Da Rold.** 2009. "The Linguistic Stratification in the Cambridge DD Copy of Chaucer's Canterbury Tales." *Neuphilologische Mitteilungen* 110: 283–297.
- TRANSKRIBUS.** 2020. Accessed September 30. <https://transkribus.eu/Transkribus/>.
- Traube, Ludwig.** 1902. *Hieronymi Chronicorum Codicis Floriacensis Fragmenta Leidensia, Parisiana, Vaticana Phototypice Edita*. Leiden: Sijthoff.
- Voigts, Linda Ehrsam.** 1989. "The Character of the Carecter: Ambiguous Sigils in Scientific and Medical Texts." In *Latin and Vernacular: Studies in Late-Medieval Texts and Manuscripts*, edited by Alastair J. Minnis. Bury St. Edmunds: D.S. Brewer.
- Walsh, John A., and Wallace E. Hooper.** 2012. "The Liberty of Invention: Alchemical Discourse and Information Technology Standardization." *Literary and Linguistic Computing* 27(1): 55–79. DOI: <https://doi.org/10.1093/lc/fqr038>
- Wikimedia Commons.** 2020. File:Queen of hearts2.svg. Accessed September 30, 2020. https://commons.wikimedia.org/wiki/File:Queen_of_hearts2.svg.
- Wittberger-Markwardt, Vreni.** 2018. "Kürzungen in Althochdeutschen Sachglossaren. Typologie der Bezeichneten Kürzungsverfahren in der Volkssprache." In *Handschriften als Quellen der Sprach- und Kulturwissenschaft: Aktuelle Fragestellungen – Methoden – Probleme*, edited by Anette Kremer, Vincenz Schwab, 61–82. Bamberg: University of Bamberg Press.
- Wright, Laura.** 2002. "Standard English and the Lexicon: Why So Many Different Spellings?" In *Language Change: The Interplay of Internal, External and Extra-linguistic Factors*, edited by Mari C. Jones and Edith Esch, 181–200. Berlin: De Gruyter.

- . 2011. "On Variation in Medieval Mixed-Language Business Writing." In *Code-Switching in Early English*, edited by Herbert Schendl and Laura Wright, 191–218. Berlin: De Gruyter, pp. 191–218. DOI: <https://doi.org/10.1515/9783110253368.191>
- . 2018. "A Multilingual Approach to the History of Standard English." In *Multilingual Practices in Language History: English and Beyond*, edited by Päivi Pahta, Janne Skaffari and Laura Wright, 339–358. Berlin: De Gruyter. DOI: <https://doi.org/10.1515/9781501504945-016>

How to cite this article: Honkapohja, Alpo. 2021. "Digital Approaches to Manuscript Abbreviations: Where Are We at the Beginning of the 2020s?" *Digital Medievalist* 14(1): 1, pp. 1–40. DOI: <https://doi.org/10.16995/dm.88>

Published: 09 July 2021

Copyright: © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Digital Medievalist is a peer-reviewed open access journal published by Open Library of Humanities.

OPEN ACCESS The Open Access icon, a stylized 'O' with a person inside, representing open access.