The long history of shortening: a diachronic analysis of abbreviation practices from the fifteenth to the twenty-first century[1]

ALPO HONKAPOHJA
University of Oslo

and

IMOGEN MARCUS
Edge Hill University
(Received 11 April 2022; revised 15 July 2023)

This article investigates continuities and changes in abbreviation practices from late Middle English to twenty-first-century digital platforms. Adopting a diachronic perspective and lexicological framework, it quantitatively analyses frequency patterns across fifteenth-century memoranda, letters and administrative receipts, seventeenth-century letters and depositions, late nineteenth-century letters, early twentieth-century letters and a subcorpus of WhatsApp instant messages dating from 2018–19. It then presents analyses of the frequencies of various abbreviation forms, such as clippings, and abbreviated lexemes, such as their use for names, over time. The article then provides a qualitative analysis of these lexeme categories over the centuries, with a focus on specific examples. Major changes to overall abbreviation density across time are identified. The forms of abbreviation also go through major change, but the types of lexemes that are abbreviated stay more consistent over time. For example, abbreviations being used for closed-class function words such as *the* and *that* are dominant from the earliest data we have looked at to the present day. Overall, the study demonstrates how situating new media abbreviation practices within a historical continuum can enhance our understanding of them.

Keywords: abbreviation, lexis, writing systems, diachronic variation, computer-mediated communication

1 Introduction

Abbreviating words instead of spelling them out in full is a phenomenon found throughout the history of written communication, from ancient inscriptions carved into stone and medieval manuscripts to modern-day instant messages. Unsurprisingly, ABBREVIATION has been studied in various fields of research, including lexicology, studies into computer-mediated communication (CMC) and palaeography. However, as these different disciplines do not necessarily engage with each other, this attention has been compartmentalised, typically only focusing on a single period, text type or type

of abbreviation. The current article therefore seeks to bridge the gap between studies of earlier abbreviation practices and those present in CMC by using a corpus-based, long diachronic approach, data from the fifteenth to the twenty-first century, and a framework that examines both abbreviation types and lexis that is abbreviated. In so doing, it aims to lay the foundation for further diachronic studies of abbreviation practices.

We define 'abbreviation' as in Bieswanger (2013: 474), as a term which is used 'to refer to all strategies that result in lexical forms that are made up by fewer characters than the full form of a word or a combination of words'. Using the shorter form means that some intelligibility may be lost, but has the advantage that the message can be fitted into shorter space or written in a shorter time. This advantage is the same whether you were writing on paper in the late fifteenth century or using an electronic device in the twenty-first. The motivation to use abbreviations, i.e. to save space and time (see Petti's 'economy of space' and 'economy of time', 1977: 22), has not changed over the centuries. There are also similarities over time in relation to which lexical items are abbreviated, as well as the types of abbreviation used. Consider the following four examples, each of which represents a period in the development of the English language: twenty-first century English (1), Late Modern English (LModE) (2), Early Modern English (EModE) (3) and Middle English (ME) (4).[2]

(1) WhatsApp message (18/02/2018, 11:59)

　　Sorry **I've** not been in touch, **I've** been doing teaching **etc** in a day so far so quite pushed for
　　time when in **Bris**! **I'll** be around on **Weds** too later.

(2) A letter by Ernest Dowson (1889)

　　Thanks for yours. Also for **Chap** […] Your letter
　　depressed me - but I will try **&** live till next week. **Satdy** I go to a
　　matinée of Minnie **T.**

(3) A letter by Daniel Fleming (1660)

　　Our adversary Studholme was apprehended yesterday by **S^{r}** William
　　Carleton at ye house of one Sturdy a Quaker there is a
　　dangerous accusation put in against him, as to ye life of his **Ma^{ty}**
　　where I found **Dr** Carleton his **Maties** chaplaine, **&c**

(4) A receipt (1462)

　　þ^{e} Fryday next **be for̄** þ^{e} fest of Seynt **Clem̄t** þ^{e} yer̄ of **kyn^{o}** Edward
　　þ^{e} ıııȷ þ^{e} íȷ · **y^{s}** resseyvyd by þ^{e} **handʄ** of **Stepħ.**

---

[2]　We start our study with late Middle English, even though many of the ME practices descend from OE and Medieval Latin. However, we have chosen to limit our focus for two reasons. Firstly, the dissemination of Latin abbreviations into vernaculars has been covered in palaeographical handbooks. Secondly, the greater survival of ephemeral text types such as receipts, memoranda and letters (as original copies on paper rather than copied into manuscripts) towards the end of the Middle Ages provides a degree of continuity for our investigation. Continuity with Old English is impossible to achieve, since Old English lacks the more informal text types.

Abbreviating expressions of time are found in examples (1) (*Weds* 'Wednesday'), (2) (*Satd* 'Saturday') and (4) (*yeȓ of kynᵍ Edward þᵉ ɯɯ* 'year of king Edward IV'). Personal names are abbreviated in (2) (*Minnie T.*) and (4) (*Stepħ* 'Stephen, Stephanus'), titles in (3) (*Sᵣ* 'Sir', *Mᵃᵗʸ* 'Majesty', *Dr* 'Doctor') and (4) (*kynᵍ* 'king') and function words in (1) (*I've* 'I have', *I'll* 'I will') and (4) (*yˢ* 'this'). The discourse-organising abbreviation 'et cetera', which indicates 'the statement refers not only to the things enumerated, but to others which may be inferred from analogy' (*Oxford English Dictionary* (*OED*)), is used in both the EModE example and the twenty-first-century one, even though the form of the abbreviation changes (*&c* versus *etc*).

However, as well as similarities over time, there are also differences. Abbreviation forms display change as well as continuity. Example (4) from the ME period contains several instances of indicating abbreviation by so-called BREVIGRAPHS, handwritten abbreviations which shorten the word by specific signs or fourishes (*be foȓ* 'before', *yeȓ*, *kynᵍ*, *Stepħ*, *handf* 'hand(e/i/y)s'). SUPERSCRIPT LETTERS are common in the EModE example (3): they include *Sᵣ*, *Maᵗʸ*, *Maᵗⁱᵉˢ*, as well as the non-abbreviating superscript *yᵉ* ('the').[3] Present-Day English (PDE) and LModE examples (1) and (2) favour abbreviating words by leaving out letters from the end by CLIPPING (*Bris* 'Bristol', *Chap* 'Chapter') or from the middle by CONTRACTION (*Weds*, *Satdy*), although the latter is also attested in the EModE example (3) (*Dr*), and in the ME example (4) (*Cleꝫt*, 'Clement, Clementinus'), albeit accompanied by a horizontal bar indicating the presence of an abbreviation. The purpose of the present study is to therefore to examine both abbreviation forms and abbreviated lexemes in order to provide a quantitative description of what exactly changes and what stays the same over time.

In order to assess both continuity and change in abbreviation usage, the study quantitatively analyses overall abbreviation frequency patterns across four periods in the history of English: twenty-first century, LModE, EModE and ME. We collected a dataset of abbreviated spellings in each subcorpus and annotated them for both ABBREVIATION FORM (e.g. brevigraph, contraction, clipping, superscript) and LEXEME CATEGORY (e.g. name, title, function word, expression of time). This dataset was then subjected to exploratory quantitative analyses, including descriptive statistics (specifically log likelihood tests). The research questions of the study are:

  I How frequent is abbreviation in each historical period?
 II What forms of abbreviation are used in each period, and how common are they?
III What kind of lexemes get abbreviated in different periods?

The quantitative analysis was then combined with secondary qualitative analyses, with a focus on specific examples. The results were cross-checked and contextualised using previous studies of abbreviation and dictionaries, such as the *OED* and the *Middle*

---

[3] Our definition of abbreviation excludes non-abbreviating superscripts (see section 3.1 below).

English Dictionary (*MED*), and corpora, such as the *Corpus of Contemporary American English* (COCA, Davies 2008–).

The article is structured as follows. Section 2 presents an overview of secondary literature across relevant disciplines. Section 3 outlines the analytical categories used to annotate abbreviation forms (3.1) and lexemes (3.2). Section 4 details the corpora used in the present study, highlights the reasons used for selecting texts and outlines our methodology. Section 5 presents the analysis, starting with overall frequencies of abbreviated lexemes in each period (5.1). It then presents quantitative, diachronic analyses of the frequencies of various abbreviation forms (5.2) and what kind of lexemes they are abbreviating (5.3). We discuss in detail names (5.4), discursive and metadiscursive abbreviations (5.5) and expressions of time (5.6), as well as new lexeme categories in LModE and twenty-first-century English (5.7). Section 6 provides some discussion and concluding remarks.

2   Research context

Abbreviation has received attention across a range of disciplines. Writing manuals often comment on the topic, giving stylistic recommendations on when to use abbreviations and when to avoid them (see e.g. Williams 1981; Lang 2019). In lexicology, the focus has predominantly been on ACRONYMS AND INITIALISMS (see e.g. Cannon 1989; Rodríguez & Cannon 1994; Klymenko 2019) and clipping (Bauer 1993; Plag 2003; Stockwell & Minkova 2003; Bassac 2004; Katamba 2005; Jamet 2009) in relation to word-formation processes. Most analyses of abbreviation are part of more general studies of lexis and word formation, in which clipping is sometimes described as a 'marginal' word-formation process, as it produces a fairly limited number of words (see Bassac 2004: 170; Jamet 2009: 20), although Rodríguez & Cannon (1994: 270–1) put forward the case for the importance of initialisms in relation to dynamic word formation. The situation is rather different in CMC studies, where abbreviation received attention at the turn of the millennium.

2.1   Abbreviation in CMC

Abbreviations have often been characterised as a feature of CMC and can indeed be found in, for example, internet relay chat (IRC), online forums, instant messaging (IM), text messaging and weblogs (see e.g. Drouin & Davis 2009; Bieswanger 2013: 463; Lyddy *et al.* 2014). The presence of abbreviation in new media was especially highlighted by what Androutsopoulos (2006: 420) calls the 'first wave' of CMC research. In these early studies, abbreviations frequently appeared in lists of 'common features of digital writing' (Danet 2001: 17) or 'Netspeak' (Crystal 2001: 17), along with emojis and non-standard spellings.

However, later studies criticised the concept of a single 'Netspeak', or internet language, put forward by first-wave CMC researchers. They suggested it was based on overgeneralisation (see Bieswanger 2013: 465–6). Quantitative studies that sought to

verify Crystal's (2001: 84) claims on the high frequency of abbreviations in 'Netspeak' never found abbreviation frequencies higher than 20 per cent (see Bieswanger 2013: 476). Furthermore, whilst early studies (see e.g. Herring 2005) found abbreviation to be characteristic of synchronous CMC modes like chat and IM, the binary distinction between synchronous and asynchronous CMC modes has more recently been found by Bieswanger (2016: 297) to be 'conceptually problematic' and 'not necessarily closely linked to actual patterns of language use'.

More recent CMC studies have looked at the relationship between abbreviation and standardisation (and the notion of post-standardisation) (see e.g. Shortis 2016 in relation to SMS text messaging). More recently still, researchers such as Zelenkauskaite (2017), Klymenko (2019) and Eberl (2020) have investigated how abbreviation usage is affected by technological length constraints on social media platforms such as Facebook and Twitter, while Heuman (2021) has investigated it in relation to the constraints and possibilities of digital writing more generally.

2.2 Abbreviation in palaeography

Medieval abbreviation has predominantly been studied within the field of palaeography because abbreviations are one of the main features used to date and localise samples of medieval handwriting, as well as identify scribes (Ker 1960: 54; Lowe 2006: 134; Kjeldsen 2013: 391, 404; Kestemont 2015: 172–3). Quantitative palaeographical studies of medieval Latin have uncovered abbreviation frequencies above 50 per cent (Honkapohja & Liira 2020), which is considerably higher than numbers uncovered by quantitative studies that sought to verify Crystal's (2001: 84) claim about the high frequency of abbreviations in CMC 'Netspeak'.

2.3 Lack of long diachrony in abbreviation studies

Both lexicological and CMC studies tend to focus on fairly recent developments in language and abbreviation types that are characteristic of the present day. First-wave CMC researchers sometimes acknowledge the use of abbreviations before computing, but do not discuss them in detail. For instance, Danet mentions that 'even private emails were often sprinkled with abbreviations some of which were already in use before computers' (2001: 18), naming *FYI* 'for your information' as a pre-computing example (ibid.). More recently, McCulloch (2019: 10–11) mentions Roman *SPQR* 'SenatusQue Populus Romanus [Senate and the Roman People]', medieval scribes using symbols such as *&* and *%*, and Renaissance scholars' Latinate abbreviations like *e.g.* and *ibid.*, before saying 'that the true golden era of acronyms began surprisingly recently' (2019: 10) and describing their rise in popularity starting with World War II (see also Cannon 1989: 99; Rodríguez & Cannon 1994). There is thus an awareness that abbreviation did exist long before computing, but it is rarely investigated.

The need for a study of abbreviation in long diachrony is highlighted by studies such as Elspaß (2002), which have compared CMC to nineteenth-century letters, but not earlier

data. As Androutsopoulos (2011: 150) notes, 'researchers who compare CMC to earlier vernacular writing, ranging from nineteenth century private letters to contemporary popular culture ([…] Bergs 2009; Elspaß 2002; […]), conclude that the novelty of digital writing is often exaggerated or lacks historical depth'. The purpose of the present study is to provide this depth. It complements the earlier study by Rodríguez & Cannon (1994), which is one of the few to attempt to cover abbreviation in long diachrony. However, we are updating it with reference to the internet and by using an empirical quantitative approach.

3   Definition of abbreviation

3.1   Form categories

As noted above, we define 'abbreviation' as in Bieswanger (2013: 474), as a neutral term which is used 'to refer to all strategies that result in lexical forms that are made up by fewer characters than the full form of a word or a combination of words'. This definition implies two things: first, the existence of a full form of the word that is abbreviated and second, that the abbreviated form of the word is shorter than the full form. Our definition includes abbreviations which represent spoken language (*'tis*, *can't*) if they result in a spelling that is shorter than the full form.

The requirement of a full form excludes words formed by clipping in which the full form is no longer transparent, such as twenty-first-century *bus < omnibus* or *pram < perambulator*. Applying this criterion to historical data is, as always, difficult, as we do not have access to the linguistic knowledge of historical informants. For example, if we encounter the word *bus* in the LModE corpus, how can we be sure if they understood it as an abbreviation of *omnibus* or a lexeme of its own? In practice, we cross-checked the forms in the *OED* and the *MED* for contextualisation. For the EModE and ME corpora used in this study (see section 4 below) we relied on the editorial decisions of their compilers.

We break down our choices in the following subsections, referring to categories used in lexicology, palaeography and studies of CMC. An important reference point here is the article by Rodríguez & Cannon (1994), who also adopt palaeographical terminology for describing early abbreviation practices.

A. Acronyms and initialisms

Category A includes words abbreviated to a single letter, including single words (5) and combinations of several words, as in (6) and (7). As it is impossible to know for sure how an abbreviation was pronounced in earlier periods, we do not distinguish between acronyms, pronounced as one word (7) and initialisms pronounced letter by letter (6). We include both forms in which the presence of abbreviation is indicated by a full stop (*v.*, *P.S.*) and ones in which it is not (*omg*, *lol*).

(5) *v.* 'very'

(6) *omg* 'oh my god', *pm* 'post meridiem', *p.s.* 'post scriptum'

(7) *lol* 'laughing out loud'

B. Clipping

Category B includes words abbreviated by deleting the end of the word (8).

(8) *Feb* 'February', *Ric* 'Richard, Ricardus', *Tho. Moore* 'Thomas Moore', *min* 'minute', *fab* 'fabulous'

C. Contraction

Category C includes words abbreviated by leaving out letters in the middle (9). We also categorise ordinal numerals, when they are indicated by a Hindu-Arabic numeral and final letters, as contractions (10). We include medieval abbreviations, in which a horizontal bar indicates the word is abbreviated, similarly to a full stop (11).

(9) *lres* 'lettres', *Mr* 'Master, mister, magister', *bday* 'birthday', *tmrw* 'tomorrow'

(10) *2nd* 'second', *9th* 'ninth'

(11) *Rīc̄us* 'Ric(ard)us'

D. Elisions and informal spellings

Form category D is more complicated to define than categories A to C, which are based on where the word is shortened. We use the term ELISION to refer to a convention in which unstressed syllables are replaced by an apostrophe, sometimes also involving blending of two words together (12). Petti ([1977](#): 25) describes it as 'silencing of letters for metrical necessity, euphony or colloquial convenience'. In phonetics, elision refers to a process in which phonemes are dropped in certain contexts, for example in medial position in unstressed syllables. Diachronically, elision is connected to the grammaticalisation of auxiliaries such as *have*, *will* or *shall*, which causes the auxiliary to be unstressed and elided. As these elided spellings make the word shorter, they fall under our definition of abbreviation.

   Words were categorised as D elisions rather than C contractions, when they were indicated by an apostrophe or there was a good reason to assume the shortened spelling was used to mimic spoken utterance. We also include informal spellings, which mimic spoken language (13), ME spellings in which the definite article appears merged with the noun head (14)[4] and French elisions (15).

(12) *'tis* 'it is', *don't* 'do not', *Edinboro'* 'Edinburgh'

(13) *wanna* 'want to', *deffo* 'definitely'

(14) *tharchaungełł* 'the archangel', *thvnuᵍsıtıe* 'the university'

(15) *c'est* 'it is', *l'oeuvre* 'the work'

---

[4] In Middle English, elision is not indicated by an apostrophe. However, spellings such as the ones given in (14) reflect the same phenomenon. These spelling practices have a parallel in Anglo-Norman and Middle French, in which spellings such as *lez roys de fraunce et dangleter* 'the kings of France and England' (Cotton Julius B I f 93v) and *sergeant darmes* 'sergeant of arms' (Cotton MS Julius C IV f. 84r) are common in the fifteenth century.

E. Logograms and self-standing symbols

Category E includes special signs (16), which are called logograms (see Honkapohja 2021: §10–§14). We only include what can be termed 'pure' logograms (LAEME §2.2.1). If the special sign is combined with alphabetic characters (19), it is classified under category F.

(16) @ 'at', & 'and', £ 'British pound', *p* 'per'

F. Brevigraphs

Category F includes the characteristically medieval abbreviations known as brevigraphs, which developed over centuries 'to shorten the labour of writing Latin' (Hector 1958: 37). Most brevigraphs originated in Latin, but were also applied to the vernacular, albeit with lower frequency (Honkapohja & Liira 2020). Brevigraphs involve replacing a commonly occurring letter combination by a single character, stroke or flourish. This made the word shorter to write than the full form, which required more pen strokes. The abbreviated part often corresponds with morphemes ((17), *bestf* 'best(is) [beasts]', the plural marker) or syllables ((18), *psones* '(per)sones'), making the category overlap with syllabic writing systems such as Japanese *hiragana* and *katakana*, although not all brevigraphs correspond to syllables.

The term brevigraph is somewhat fuzzy and may be divided into subcategories (see Honkapohja 2013: table 1). For example, Cappelli (1899) and Hector (1958), make a distinction between 'special signs of abbreviation', which correspond with specific graphs, indicating that two or three letters need to be supplied, and 'general signs of abbreviation', which simply indicate that the word has been abbreviated. The very common 'macron' or 'title', a horizontal bar written above letters, is often used to abbreviate a nasal (as in *im̄ediately* 'im(m)ediately'), where it is used as a special sign, but can also simply indicate that the word has been abbreviated (as in *Rīcus* 'Ric(ard) us' in (11) above, where it is used as a general sign).

If the macron was used to indicate a nasal, we classified it as F brevigraph. If it was used as a general sign of abbreviation, we classified it as B clipping (*Thom̄* as in 'Thomas') or C contraction (11) based on what part of the word was abbreviated. The latter usage is analogous to modern full stop – a multipurpose sign, which can be used as general sign of abbreviation, but also has other uses, including in punctuation or to indicate decimal points in numerals (see (5) and (6)).

Another point of overlap is with category E logograms and self-standing symbols, since brevigraphs could sometimes be used independently. For example, the 'crossed-p' *p* 'per [through, during, by means of]' was found in our data both independently to stand for the Latin preposition *per* (16) and to abbreviate the syllable as a part of a word (*psones* '(per)sons', (18)). In these cases, we used the following criteria: if the special graph was self-standing, we classified it as E; if it was used as part of a word, we classified it as F.

(17) *xp̄ofeɼ* '(Christo)pher', *bestf* 'best(is)[beasts]'
(18) *souꝯaɪn* 'sov(er)aign', *psones* '(per)sones'

(19) *Xmas* '(Christ)mas'

G. *Superscript letters*

Category G includes abbreviations in which part of the abbreviated word, often the last letter, is written above the line (20). Following our definition of abbreviation, we have not included non-abbreviating superscripts, such as $þ^e$, y^e 'the', which are fairly common in the EModE and ME periods (see (3) and (4) above).

(20) w^t 'with', m^{tie} 'majesty'

3.2 *Lexeme categories for the purpose of function analysis*

After we had categorised the abbreviated words based on the form of abbreviation, we turned our attention to which lexemes were being abbreviated, using an exploratory bottom-up methodology. We sorted these lexemes into various categories (see table 1).

Some of these lexemes represent clear semantic fields, such as words referring to date and time or personal names. Others require further explication.

Some of the lexemes which are frequently abbreviated are part of address formula in correspondence (see e.g. Nevala 2004). As Hasenohr (2002: 80) notes, words that get abbreviated are ones 'which come often under the pen'. The conventions of addressing people are a good example of this, as 'people relied on the various letter-writing customs which developed over the centuries' (Nevala 2004: 2127). Another group of

Table 1. *Lexeme categories*

Lexeme category	Examples
Date/time	Dec, xmas, Sat
Names (personal and place names)	Jo, Tho. 'Thomas', Bydenħm
Titles	Mr, Rev., G.P.
Address/affection/honorific terms	Mdear, darl, souᵒaın, m^{tie}
Discursive and metadiscursive items	i.e., etc., P.S., re:, e.g., OK
Function words	tho, btwn, &, w^t
Measurements	℔ 'libra/pound', £ 'British pound', 2^d 'two pence', %
Other lexical words	vestmentſ 'vestiments', ptıe 'party', ımpⁱsonment 'imprisonment', inclos'd 'enclosed', Govt. 'government', email, vocab 'vocabulary', tix 'tickets'
Ordinal numerals	vıȷth 'seventh', 26th, 3rd
Exclamations	wtf, omg, lol
Brand and organisational names	M&S 'Marks & Spencer'

lexemes which commonly get abbreviated are function words, including prepositions, pronouns and closed-class adverbs (Biber *et al.* 2000: 2.2.3.1).

Lexemes which are abbreviated reasonably frequently are what we call discursive and metadiscursive items, following Álvarez-Gil & Bondi (2021: 2). These devices do not have to do with the propositional content of a text, but rather 'the metadiscursive level, which in principle provides the audience with material guiding them through their process of interpretation' (Álvarez-Gil & Bondi 2021: 2, quoting Williams 1981: 121–2). We included in this category such commonly used and abbreviated devices as *etc.* and *P.S.* and also *Ok*. The latter is mainly discursive, because in text types like IM it is used to express 'assent, concession or approval, esp. with regard to a previous statement or question' (*OED*, s.v. *okay*, def. B1).

The initial lexeme categories, such as date/time, function words and address terms, relevant to the medieval data, were also sufficient for EModE and LModE corpora. However, there were some categories that only emerged when we examined the data from the later periods: exclamations such as the internet staples *omg*[5] and *lol*, and brand and organisational names. Finally, there were a number of miscellaneous lexical words across various semantic fields, which were gathered under the umbrella term 'other lexical words'.

4 Data and methodology

The study was designed to provide comparative written data from the twenty-first century, LModE, EModE and ME. As mentioned in section 2.3 above, the starting point was that claims made about the novelty of digital writing tend to lack historical depth (see Androutsopoulos 2011: 150). Our digital writing data were a subcorpus of twenty-first-century WhatsApp instant messages, dating from 2014–19, collected and compiled for the *Transhistorical Corpus of Written English* (TCWE, Marcus & Maden-Weinberger 2021). The IM data were collected from people in their teens, twenties, thirties, forties, fifties and sixties. It contains a mixture of both one-to-one and group chats.

A key reason instant messaging (IM) was chosen as the digital text type under investigation is the existence of notable previous research done into the use of abbreviation within it (e.g. Tagliamonte & Denis 2008). Tagliamonte & Denis (2008: 3) conceptualise IM a 'unique new hybrid register, exhibiting a fusion of the full range of variants from the speech community: formal, informal, and highly vernacular'. Furthermore, like other kinds of CMC, IM has been conceived as a 'blend or hybrid of written and spoken aspects of language' (Androutsopoulos 2011: 5). In light of these characterisations, we aimed to select historical written data which either represented informal communication, would have been written down quickly enough to match the partly synchronous nature of IM, or which had an element of 'spokenness' (see Amador-Moreno 2019: 68).

---

[5] The *OED* entry for *OMG* (*OED*, s.v. *OMG*, def. A) includes a citation dating from 1917, although all the other citations date from the mid 1990s onwards.

We chose epistolary data from all three of the earlier periods because the 'spokenness of private correspondence' (Amador-Moreno 2019: 68) is a characteristic noted by a number of scholars (see e.g. Biber & Finegan 1989: 515; Fitzmaurice 2002: 19–28, who described 'the practice of letter writing as conversation'; Palander-Collin 2010; Rutten and van der Wal 2014; Marcus 2017, among others). The letters have been taken from *The Corpus of Early English Correspondence* (CEEC, Nevalainen *et al.* 1998) and the *Corpus of Late Modern English Prose* (Denison 1994), a corpus of informal private letters by British writers dating from 1861 to 1919. We also took a selection of the early twentieth-century *Upton Letters* from Project Gutenberg and a selection of transcripts of World War I *Letters to Loved Ones* from the Imperial War Museums website. To provide a point of comparison with IM, we specifically chose private personal letters rather than public, business letters, because the former usually contain more informal, everyday discourse than the latter.

For the two earlier periods, we complemented correspondence with other text types containing everyday, non-literary discourse with 'speech-like' qualities, including court records which recorded spoken-language testimonies given in court. The Early Modern Corpus therefore also contains witness depositions from the *English Witness Depositions 1560–1760: An Electronic Text Edition* (ETED, Kytö *et al.* 2011). They were chosen because, according to Culpeper & Kytö (2010: 18), they are firmly within the 'speech-based' category. Using CEEC was also limited by the availability of data, since like many historical corpora, it is based on printed editions some of which silently expand abbreviations.

For ME, we used the *Middle English Local Documents* (MELD) corpus, which is based directly on manuscript sources and encodes abbreviations. We picked two text types in addition to correspondence: statements, specifically receipts, and memoranda. Statements 'may be made by a person or group, usually in first person but sometimes in the third' (see Stenroos *et al.* 2020: 51). A particularly common form of statement from the medieval period is the receipt, which is 'a brief acknowledgement of the receipt of money, legal documents, goods, - or prisoners' (2020: 51). They are often written in the hand of the person who signs them. Memoranda are a very heterogeneous and variable group but are usefully classified by Stenroos, Bergstrøm & Thengs as 'records of various kinds', which 'note down facts, events or decisions' (2020: 50). They usually contain 'no formulaic content apart from an initial Memorandum' (2020: 50–1).

It should also be noted that the selection of corpora is limited by the availability of texts containing abbreviations, since an editorial practice of expanding them silently is commonplace. Historical linguistic corpora often inherit the practice of silently expanding abbreviations from printed editions. In some born-digital corpora abbreviations are not deemed to be important for research questions and are expanded for this reason (see Honkapohja 2021: §3–§7). MELD and ETED are compiled directly from manuscripts and encode abbreviations. This also meant that we could follow the editorial decisions of their compilers on what constitutes an abbreviation. For CEEC, we used the letters which encode abbreviations rather than expanding them. An overview of data sources and word counts is presented in table 2.

Table 2. *Overview of corpora and subcorpora*

Language time period	Subcorpus	Text type	Word count	Total word count for time period	Total word count of whole corpus
ME (1066–1500)	*Middle English Local Documents Corpus* (MELD)	15th-century letters	3,323	16,486	
	Middle English Local Documents Corpus (MELD)	15th-century statements, receipts	1,705		
	Middle English Local Documents Corpus (MELD)	15th-century memoranda	11,458		
EModE (1500–1700)	*The Corpus of Early English Correspondence* (CEEC)	17th-century letters	21,580	25,871	
	English Witness Depositions 1560–1760: An Electronic Text Edition (ETED)	17th-century depositions	4,291		
LModE (1700–1945)	*The Corpus of Early English Correspondence* (CEEC)	Late 19th-century letters	14,456	25,883	
	Corpus of Late Modern English Prose, Project Gutenberg, Imperial War Museums Website	Early 20th-century letters	11,427		
21st-century English (2000–present day)	*Transhistorical Corpus of Written English* (TCWE)	21st-century instant messages	21,228	21,228	
					89,468 words

Table 3. *Dataset*

Abbreviation	Expanded form	Form	Lexeme category	Language
Tho.	Thomas	B. Clipping	Name	Eng
tho	though	B. Clipping	Function word	Eng
P.S.	post scriptum	A. Acronym	Discursive and metadiscursive item	Lat
WTF	What the fuck	A. Acronym	Exclamation	Eng

To analyse the data we located abbreviations manually and annotated them in a dataset, which formed the basis for the quantitative corpus-based analysis. Table 3 illustrates what the dataset looks like. For instance, *Tho.* is a clipping of the English name 'Thomas', which is why it is listed as clipping under 'Form', 'Name' under Lexeme category and 'Eng' under Language.

5   Results

5.1   *Overall frequencies of abbreviation*

Our results indicate a high frequency of abbreviation usage in the medieval period: a mean of 140.08 abbreviations per 1,000 words (14 per cent of all words in the ME data being abbreviated), which compares with lower densities in the EModE and LModE periods of 82.82 abbreviations per 1,000 words (8.8 per cent of the total word count) and 34.07 abbreviations per 1,000 words (3.4 per cent of the total word count) respectively. After this steady decline there is a clear increase in the twenty-first century (57.47 instances per 1,000 words, which translates to 5.7 per cent of the total number of words in this subcorpus) (see figure 1).

A log likelihood test (on raw word counts) was conducted to analyse the frequency changes between the four time periods studied. From ME to EModE there was a drop of 36.58 per cent, from EModE to LModE a drop of 56.86 per cent, and then from LModE to twenty-first-century English, a major increase in relative frequency of 60.47 per cent (%DIFF), with a high level of confidence. The log likelihood score for this last comparison is 119.27, and the Bayesian Information Criterion score is 108.51, which, coupled with the 60.47 per cent difference in effect size, makes it statistically highly significant. There is also a large effect size difference of 155.55 per cent between nineteenth- and twentieth-century letters, suggesting that abbreviation had become very infrequent at the turn of the twentieth century, before an even larger increase from the twentieth-century corpus to the twenty-first century (effect size 199.87 per cent).

The differences between text types within each time period are mostly smaller than between periods. The difference between ME statements and memoranda is 35.53%, whilst the difference between seventeenth-century letters and depositions is 0.77 per cent. When the results presented above are broken down by the different text types in
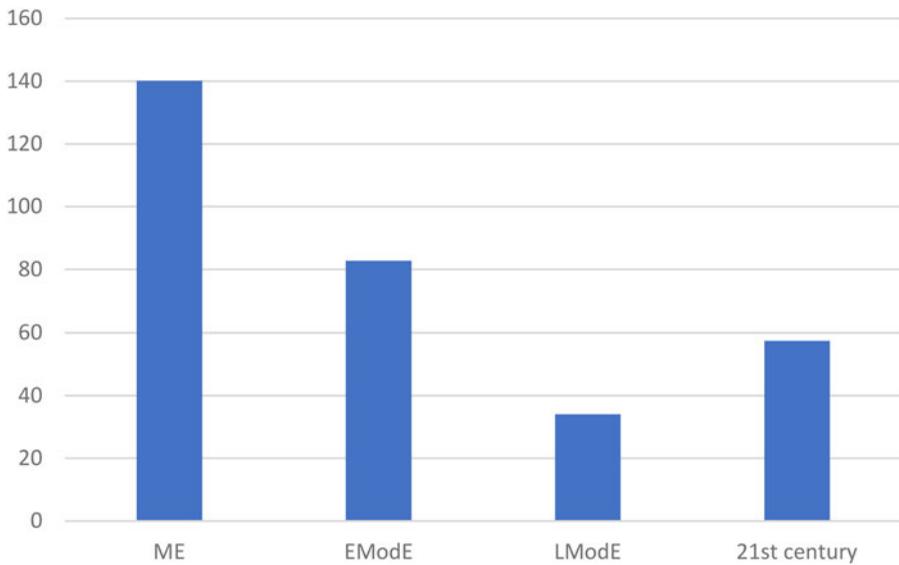
Figure 1. Abbreviation density normalised per 1,000 words, by period
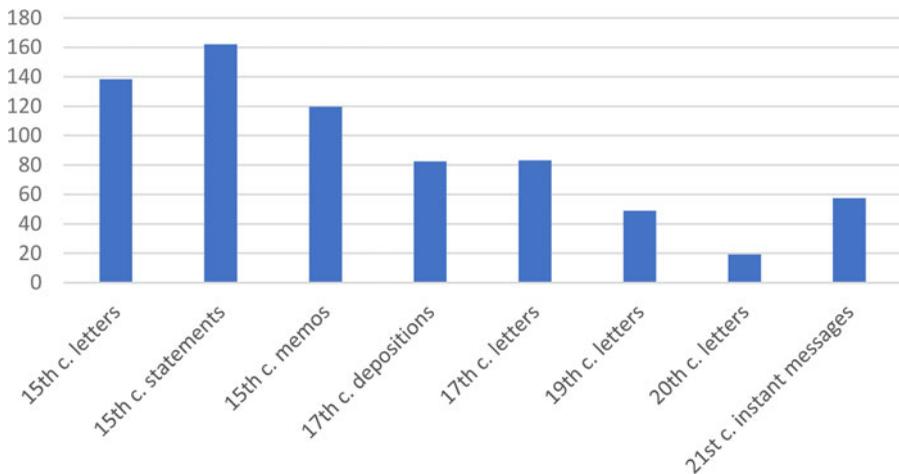
Figure 2. Abbreviation density normalised per 1,000 words, by text type

the four time periods (see figure 2), it is possible to see similar frequencies across ME letters (13.8 per cent), statements (16.2 per cent) and memoranda (12.8 per cent), and across seventeenth-century letters (8.3 per cent) and depositions (8.2 per cent). Since differences between text types are smaller than diachronic differences, this supports our decision to group them together.

5.2 Changes in abbreviation forms

Our analysis reveals clear changes in the prevalent forms of abbreviation used over the centuries across different text types, as illustrated in figure 3 and summarised in table 4.

The chart and corresponding table show a major transformation in the abbreviation landscape. The ME period shows a high incidence of brevigraphs, which drops drastically in the second period and virtually disappears in the third and fourth ones. Meanwhile, logograms and superscripts reach their peak in the EModE period and then taper off – although logograms like £, % and & are still attested in the twenty-first-century corpus, giving the normalised frequency 0.85/1,000 words. In contrast, the use of acronyms, clippings and elisions increases towards the present-day, while contractions reach their highest usage in the LModE period. Thus, our data reveal a major shift in abbreviation forms, primarily occurring between the EModE and LModE periods.

Three forms of abbreviation show an increase in the twenty-first-century corpus. They are elisions, clippings as well as acronyms and initialisms. The last category shows a consistent rise across the periods with a substantial change from 4.06/1,000 words to 7.44/1,000 from LModE to the twenty-first century. These findings support the claims made by McCulloch (2019) and Cannon (1989) that the present day is the golden era
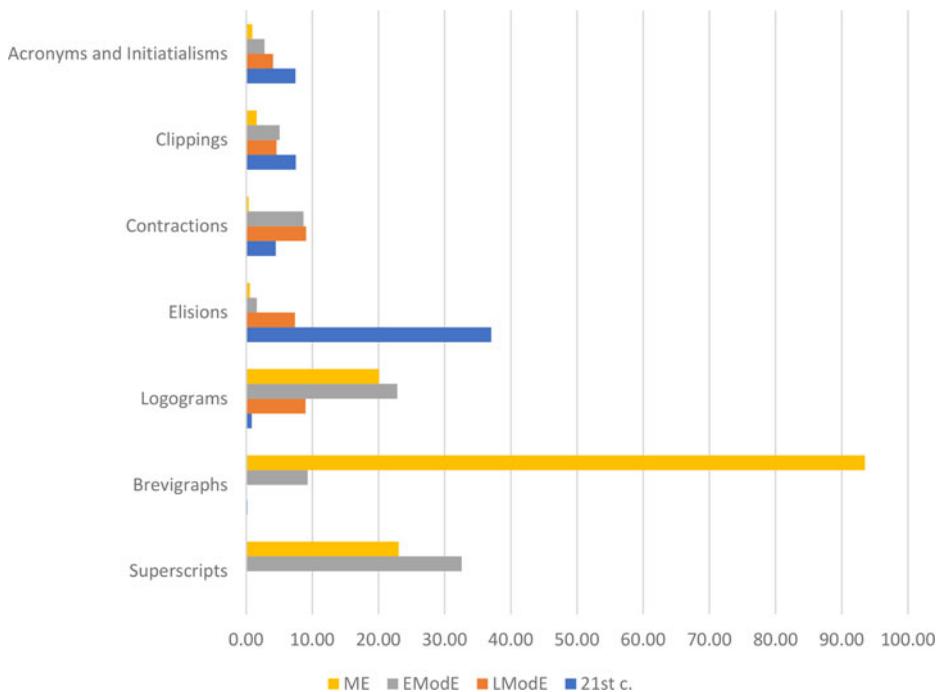

Figure 3. Numbers of abbreviation forms per 1,000 words

Table 4. *Changes in abbreviation form (normalised frequency per 1,000 words)*

	ME	EModE	LModE	21st century
Acronyms and initialisms	0.92	2.77	4.06	7.44
Clippings	1.59	5.06	4.59	7.49
Contractions	0.37	8.68	9.05	4.48
Elisions	0.56	1.62	7.36	37.03
Letter number homophones	0.00	0.00	0.00	0.00
Logograms	20.08	22.82	8.98	0.85
Brevigraphs	93.51	9.29	0.00	0.19
Superscripts	23.05	32.58	0.03	0.00

of acronyms. The increase in clippings from 4.59/1,000 to 7.49/1,000 is also unsurprising, as clipping is considered to be a productive word-formation process in twenty-first-century English (see section 2 above). Consequently, the two word-formation-related categories which have received the majority of attention in lexicology do exhibit a moderate increase.

The biggest and most sudden increase, however, takes place with elision. During the LModE period, elisions show a modest increase with a frequency of 7.36/1,000 words. In the final period, the normalised frequency of elisions skyrockets to 37.03/1,000. This increase can be partly attributed to the development of auxiliaries over time (with constructions such as *you'd* becoming more prevalent). The frequent use of elisions aligns with earlier studies of IM behaviour. For instance, Baron (2004) found that college students used elisions instead of full forms in 65.3% of instances, but her college IM corpus 'yielded surprisingly few CMC abbreviations or acronyms' (2004: 412). Our results are similar, with elisions being the most popular form of abbreviation. It is quite likely that this is due to the informal register of LModE personal letters and twenty-first-century IM.

In the final period, in addition to elisions, some informants use non-standard abbreviated spellings like *defo* or *probs*, which require conscious effort as they are not suggested by predictive text. Out of 786 elisions, only 34 are informal spellings. These spellings, along with their frequencies, are all listed in (21) below:

(21) *wanna* (3), *fabbo* (1), *tix* 'tickets' (1), *deffo* (5), *probs* (1), *ya* (3), *cos* (3), *dunno* (1), *gonna* (3), *okay* (4), *sayin* (1), *camp n furnace* (3), *nite* (1), *okies* (1), *soz* 'sorry' (1), *ta* 'thanks' (1), *u* (1)

The intentional use of non-standard spellings is usually considered in CMC research to be a way for IM users to signal 'to others that the producer understands the particulars of the linguistic characteristics of IM, and therefore, understands the IM culture' (Quan-Haase 2009: 37). However, these spellings are fairly rare in our data. What people actually use most of the time are elisions indicated by an apostrophe.

5.3 Changes in abbreviated lexeme categories

Lexeme categories exhibit both continuity and change from ME to the twenty-first century. Figure 4 provides a visual representation, while summarised results are presented in table 5.

The results show more continuity than abbreviation forms, although some notable changes can be observed. The main point of continuity is that abbreviations used for grammatical function words like *with* or *that* remain highly frequent across all four time periods. On the other hand, the biggest change is observed in the category 'Other lexical words', which shows a major fall from 48.70/1,000 words in the ME and EModE periods to 18.29/1,000 words. This change can most likely be explained by
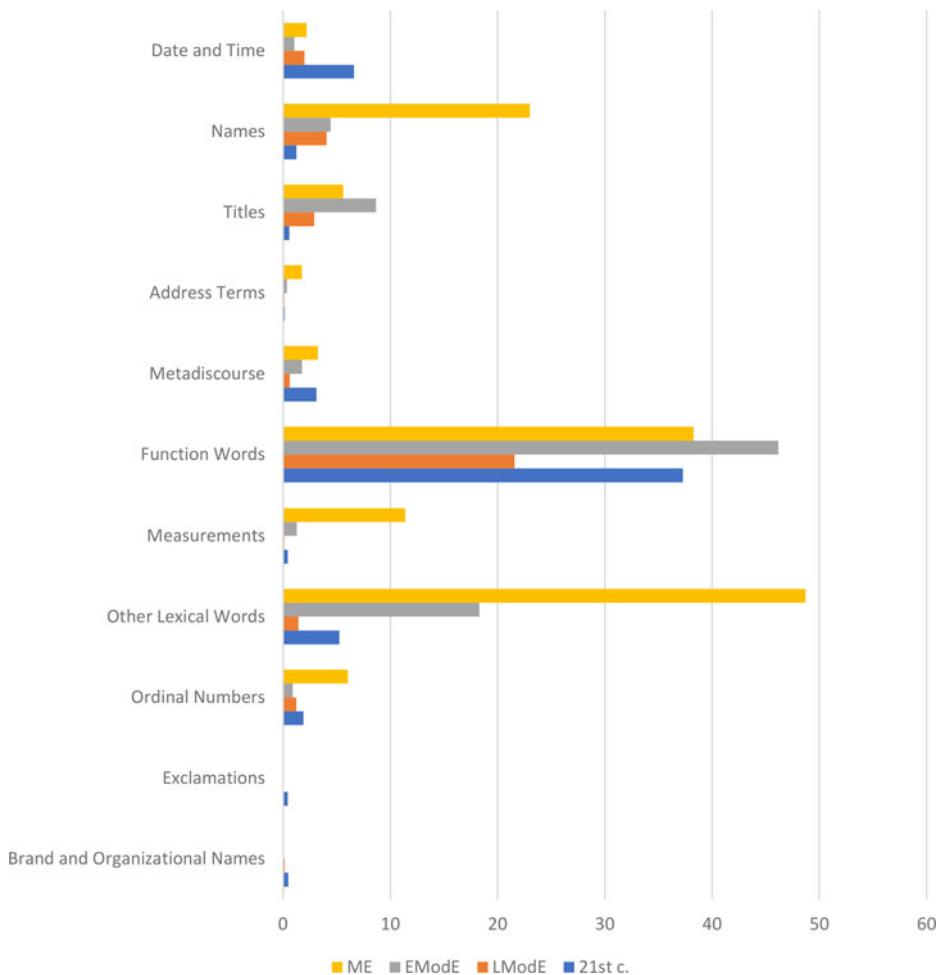

Figure 4. Frequencies of abbreviations across different lexeme categories (normalised per 1,000 words)

Table 5. *Frequencies of abbreviations across different lexeme categories (normalised per 1,000 words)*

	ME	EModE	LModE	21st century
Date and time	2.18	1.05	1.98	6.60
Names	23.00	4.43	4.04	1.22
Titles	5.58	8.65	2.88	0.57
Address terms	1.74	0.35	0.10	0.14
Discursive and metadiscursive items	3.23	1.75	0.62	3.11
Function words	38.27	46.17	21.56	37.26
Measurements	11.38	1.26	0.10	0.42
Other lexical words	48.70	18.29	1.43	5.23
Ordinal numbers	6.02	0.88	1.21	1.88
Exclamations	0	0	0	0.42
Brand and organisational names	0	0	0.14	0.47

the major change that occur in abbreviation forms. During the ME period, scribes employed brevigraphs to abbreviate all kinds of lexical words, such as *handꝛ* 'hands', *vestımentꝛ* 'vestments' or *p̄tıe* 'party' (table 1). However, after the ME period the use of brevigraphs steadily decreased, which can plausibly explain the reduced number of 'other lexical words' being abbreviated.

Six out of nine lexeme categories ('Other lexical words', 'Names', 'Address terms', 'Discursive and metadiscursive items', 'Measurements', and 'Ordinal numerals') exhibit higher abbreviation frequencies in ME compared to the subsequent periods. The high numbers in the ME period can be attributed to the higher number of abbreviations and the versatile nature of brevigraphs.

However, while abbreviation is most common in the ME corpora, there are three categories in which the normalised frequencies of abbreviations are higher in later periods. These categories are function words, titles and date and time. Two of these reach their highest peak in the subsequent period, EModE. The high frequency of abbreviated function words can be linked to the widespread use of a single abbreviation form: the superscript. The EModE period can be considered the golden age of superscript abbreviation, with many function words abbreviated using this form.

Another category which peaks in the EModE period is titles, which are, by far, most numerous in this period. The reasons for the results are hard to verify, but they may be connected to the elaborate politeness conventions and rigid social structures of early modern England (see Jucker 2020). Titles also display considerable variation between periods, with no single title found in all corpora. The abbreviation *Mr.* 'master/mister' and *Mrs.* 'mistress/missus' are not found in our data, despite the high number of personal names. The following sections break down some of the most noteworthy categories via a qualitative analysis.

5.4 Names

The category of names displays a drastic decline from ME to the subsequent periods, with the lowest number of names observed in the twenty-first century. These observations can most likely be attributed to text types and ethical considerations associated with the TCWE corpus. The high frequency of names in our ME data can be explained by the presence of lists of abbreviated witnesses' names in ME memoranda contained in a subcorpus. The EModE and LModE datasets, consisting mostly of correspondence, display very similar frequencies to each other in terms of names. In the twenty-first-century IM corpus, the low frequency of names is due to the removal of personal names from the data to maintain confidentiality (only abbreviated place names were included in the count). Hence, in this case, the changes appear to be linked to specific dataset issues rather than alternations in abbreviation practices.

There is, however, something else that takes place with abbreviations for names, which emerges from qualitative analysis. If we examine the forms of abbreviation used for names, we find abbreviations becoming more 'efficient'. In ME, personal names are primarily abbreviated by brevigraphs (22) with only regnal years sometimes being reduced to initials: rȓ 'regni Ricardi' (see table 7). In the EModE corpora, names are often clipped or contracted (23). In the LModE period, personal names are reduced to initials (24). Names of people are found abbreviated in all periods except the twenty-first, where names were omitted to protect the identity of informants, as noted above.

(22) *Roƀt* 'Robert', *wıƚƚ* 'William', *Olíuerſ* 'Oliveris'
(23) *Tho. Moore* 'Thomas Moore', *Wm* 'William', *Jo.* 'John'
(24) *P.L*, *H.B.*, *M.*

5.5 Discursive and metadiscursive abbreviations

The category of discursive and metadiscursive items shows both considerable change, but also continuity in the form of Latin abbreviations that are still used today, especially in more formal registers. Table 6 presents examples of discursive and metadiscursive abbreviations.

As table 6 shows, the only abbreviation classified as a discursive and metadiscursive item that can be found in all periods is 'et cetera'. This medieval Latin manuscript abbreviation has survived to the present day, although the form of the abbreviation has changed. Somewhere between EModE and LModE, *&c.* became *etc*.

Other medieval Latin discursive and metadiscursive abbreviations such as the symbol ꝶ 'Nota' or M^d 'memorandum', typically found in texts classified as memoranda (see Stenroos *et al.* 2020: 50–1), are not used in later periods. The EModE corpus also includes two Latin abbreviations not found in the other sub-corpora: g^o for 'ergo' and

Table 6. *Discursive and metadiscursive abbreviations*

ME	EModE	LModE	21st century
&c-	&c~	etc.	etc
ꝝ 'Nota'	–	–	–
M^d 'memorandum'	–	–	–
Jt^9 'Item'	–	–	–
–	Viz (Videlicet)	–	–
–	g^o 'ergo'	–	–
–	–	PS	P.S.
–	–	–	TBC
–	–	–	re
–	–	–	BTW
–	–	–	Ok, okay, k, okies

viz for 'videlicet'. The latter means 'that is to say; namely; to wit' (*OED*, s.v. *videlicet*, def. A) and is defined by Moore as 'the expository apposition marker *videlicet*: a discourse marker introducing an amplification, more precise explanation or specifying list (2006: 246). Both of these abbreviations were part of a Latin bibliographical system also used in the vernacular (see Stam 2017: 79; Honkapohja & Liira 2020: 283). Yet another Latin abbreviation, *PS* for 'post scriptum', is first attested in English in seventeenth-century correspondence (*OED*, s.v. *P.S.*, n. shows the first attestation in 1616).[6] However, *P.S.* is only found in our two more recent subcorpora, LModE and twenty-first-century English.

Otherwise, the IM data display a range of discursive and metadiscursive abbreviations not found in the other subcorpora, such as the clipping *re:* 'regarding', usually found in email communication, the acronyms *BTW* 'by the way' and *TBC* 'to be confirmed', and the ubiquitous *ok*, which can be abbreviated to just *k* or written as *okay*.[7] This innovation is not surprising given that, according to Androutsopoulos (2011: 150), CMC has quickly developed 'new means of textual cohesion, strategies for quoting and addressing in multi-party environments, and strategies for resolving misunderstandings with deixis'. However, our data indicate that, in IM at least, these newer abbreviations exist alongside the older, more established ones such as *etc* and *P.S.*

5.6 Expressions of time

The category of expressions of time is consistently abbreviated in all periods, but there is a noticeable increase in abbreviation frequency in the final period. Table 7 provides a breakdown of the subcategories within the time category.

---

[6] The full form *postscript*, n. is first attested earlier in a 1523 letter by Cardinal Wolsey (*OED*).

[7] Whilst the acronym *BTW* is not restricted to CMC, a search for it in COCA reveals that it is typical of it. The 6,591 instances of *BTW* in 1,001,610,938 words of COCA data from 1990 to 2019 are largely to be found in blogs (labelled as BLOG) and other websites (labelled as WEB).

Table 7. *Expressions of time*

Period	ME	EModE	LModE	21st century
Year	yᵒe, yer̃, ʒer̃, aᵒ, aᵒ dm, r̃ 'regni Ricardi'	Anno dnī, A.D., Anno Dom, Anno do:	Yrs	yrs
Month	Feb, Febrù, Aprell, Aprill, septᵒ, octob, Octobr̃, Nouembr̃, Nouembr, k 'kalendas', Menſ 'Mensis'	Jan:, Jan., Feb., Febryᵉ, Mar., Aug., Oct., Novem., Nov.	Jan., Aug., Sepr, Sept., Oct, Nov, Dec	Jan, Feb, sept, oct, dec
Day of the week	Satday	—	Thurs, Satdy, Mony	mon, tues, Wed, Thu, Thurs
Holidays	Estᵒwoke, Estᵒ, Anūcācoñ 'annunciation'	Xmas, Xtmas, Midsumēr	—	NY, NYE, nye, bday, xmas, Xmas
Time of day	—	—	p.m., a.m., hrs, o'clock, eve	pm, am, min, mins, morn, tmrw, o'clock

In all corpora, abbreviated forms meaning 'year' are observed. There are, however, some variations in the specific abbreviations used across different periods. Latin datings by regnal years are exclusive to the ME corpus, while abbreviations for 'Anno Domini' (*aᵒ dm*, *A.D.*, *Anno do:*, *Anno dnī*, *Anno Dom*) are present in both ME and EModE corpora. During the EModE period, 'Anno Domini' is the sole abbreviation for denoting years and is used in both correspondence and witness depositions. In both LModE and twenty-first-century corpora, the abbreviation for year is *yr*. Therefore, the lexeme for 'year' is abbreviated in all periods, and the same applies to months.

Names of months are found abbreviated in all periods with reasonably high frequencies. The only month to be abbreviated in all four is 'October', but other months are also abbreviated. Much like in personal names, there is a tendency towards more efficient abbreviation. Abbreviations become shorter, with clipping becoming the norm (e.g. *Jan.*, *Feb.*, *Oct.*, *Nov.*). Clippings are already prevalent in the EModE corpora. However, LModE letters still display some variation (for example, 'September' is abbreviated both as the clipping *Sept.* and the contraction *Sepr*).

References to specific times of the day become more common in the later corpora. Abbreviations related to hours of the day (*o'clock* and *A.M.* and *P.M.*) are only found in the last two corpora. Abbreviation for 'minutes' is exclusive to the twenty-first-century corpus. The higher incidence of more precise expressions of time in the twenty-first century can partly be attributed to the synchronous nature of IM – keeping in mind that the medium is defined partly by its synchronicity (Tagliamonte & Denis

2008; Androutsopoulos 2011: 5) – when using IM, people often discuss meeting at a certain time.

5.7   New categories in LModE and twenty-first-century English

Two lexeme categories were completely missing from ME and EModE, namely the use of exclamations and brand names. Both of these are, however, fairly low-frequency phenomena. There were only nine tokens which were classified as exclamations. There are three examples of *lol* 'laughing out loud', two of *omg* 'oh my God', one of *tbf* 'to be fair' and three of *O* 'oh' in the 212,228 words of the IM data (although 'O' is only arguably an abbreviation, it does fall under our definition of an abbreviation by being shorter than its variant spelling *oh*).

The numbers of these last four forms may superficially appear low, given that they are often impressionistically seen to be characteristic of CMC. However, cross-checking the abbreviation in COCA and secondary literature suggests similar developments. The frequency of *lol* in COCA for example, is 11,009 instances in 1,001,610,938 words of data, from 1990 to 2019, representing 0.001 per cent of the total word count for that twenty-nine-year time period. Similarly, in a previous study of IM, Tagliamonte & Denis (2008: 11–12) found *haha* to be a much more frequently used form to symbolise laughter (16,183 instances of *haha*, comprising a 1.47 per cent proportion of the total IM word count compared to 4,506 instances of *lol*, comprising 0.41 per cent of the total word count) (see also Schneebeli 2020: §1). They therefore note the 'sheer infrequency of the so-called "characteristic IM forms"' (2008: 12), including *lol*, in their data, and state that they are, according to their analysis at least, 'much rarer than the media have led us to believe' (2008: 12). It is also worth noting that our study only includes abbreviation and does not take into account laughing emojis (😅 😂 😆 🤣 and even, more recently, 💀).

Only two abbreviations were classified as 'Brand and organisational names' in the nineteenth-century letters subcorpus: *Archëol. Soc.*, which stands for 'Archeological Society' and *F.O.* 'Foreign Office'. The final period, on the other hand, contains thirteen tokens classified in this category: *M&S* 'Marks and Spencer' (2), *U3A* 'University of the Third Age' (1), *FB* 'Facebook' (1), *LIMF* 'Liverpool International Music Festival' (2), *camp n furnace* 'Camp and Furnace' (3), *C&F* 'Camp and Furnace' (1), *BL* 'British Library' (3).

The importance of acronyms in the category of brand and organisational names is shown by the fact that all the abbreviations in this category, with the exception of the elision/informal spelling in *camp n furnace*, are acronyms. Even *Camp n Furnace*, a nightclub in Liverpool, is once abbreviated to the acronym *C&F* – a usage which would be familiar to all chat participants. Therefore, despite not being very frequent, the uses of acronyms and initialisms that are present in our data arguably add to our understanding of acronym and initialism as productive, dynamic processes of word formation in English.

6 Discussion and concluding remarks

This study aimed to bridge the gap between medieval abbreviation practices and those found in CMC by employing a corpus-based, long diachronic approach, complemented by qualitative analysis. The results shed light on major changes in both the overall frequency of abbreviation and popular abbreviations used in each period. Our study demonstrates that, far from being specific to CMC, abbreviation was more common when all texts were handwritten. In the medieval data, approximately 14 per cent of all words are abbreviated, followed by a steady decline to 8.8 per cent in the EModE data, and 3.4 per cent in the LModE data. After that, there is an increase from 3.4 per cent of the total word count in LModE to 5.7 per cent in the twenty-first-century data. Consequently, while the frequency of abbreviation usage in ME and EModE texts is much higher than in both LModE and the present day, there is a statistically significant increase from LModE to the twenty-first century. This particular finding lends support to the idea that the arrival of the internet and CMC has increased the use of abbreviation since the LModE period.

Both continuity and change can be observed in relation to abbreviated lexemes over time, a complex situation which the above analysis has hopefully shed some light on. Abbreviation forms undergo major changes. A notable change in abbreviation forms is the decline in the use of brevigraphs and superscript abbreviations, which were characteristic of handwritten texts. The biggest change occurs between EModE to LModE, showing a considerable drop in frequency and a major change in abbreviation forms. Rodríguez & Cannon (1994: 265) highlight social changes which might partly explain this decrease. They note that from 'the early 17th century there was a notable reduction of abbreviations for literary use, and the bar and other brevigraphs were generally discarded. During the 18th century we still find a considerable number of abbreviations in legal documents . . . but their use was forbidden by Parliament during the reign of George II (1727–60)' (see also Hector 1958: 23, 29). In contrast, the forms of abbreviation that receive the most attention in lexicology, such as clippings, acronyms and initialisms, show an increase towards the present day. However, the most significant type of increase is the rise in the use of elision in the final period, the frequency of which eclipses all of the other developments.

Our study also highlights the surprising longevity of individual abbreviated forms, which can be considered fossilised. These include the logograms £, % and &, all of which are medieval or earlier in origin (see e.g. Cappelli 1899). Another example is the abbreviation 'Christmas' as *xmas*. This practice descends from an early Christian tradition of *Nomina Sacra* 'Holy Names', in which Greek letters were used to abbreviate words, such as *xps* for 'Christos' (see Traube 1907). In the ME corpus, this practice can be found, for example, in the name *x̄p̄ofeͬ* '(Christo)pher' (see (17) above). The application of this abbreviation to 'Christmas' is also an old practice. According to the *OED*, the earliest attestation is in Old English in the *Anglo-Saxon Chronicle c.* 1100 as *Xpes mǣsse* 'Christ's mass' (s.v. *X*, n.). These kinds of fossilised forms demonstrate the longevity and cultural persistence of abbreviations.

When an abbreviation fulfills a discursive or metadiscursive function, it can become extremely widely used in this capacity. Two examples of such abbreviations are *etc.* and *P.S.*, which have become so entrenched in our culture that they are frequently used even in an informal speech-like text type such as IM. Another more recent example is *ok*, which is 'believed to be a short form of "Oll Korrect," which in turn most likely came from a fad of "comical misspellings" in an 1838 satirical article about grammar published in Boston' (Lang 2019: 153; see also Rodríguez & Cannon 1994: 266–7 and Metcalf 2010). This American English abbreviation has subsequently achieved massive international popularity across the world's languages in a relatively short time. Interestingly, it was not yet used in the LModE corpus by any of our informants, all of whom were British.

The present study provides a broad overview of abbreviation use over a long period of time. It is important to acknowledge the limitations and weaknesses inherent in this approach. Firstly, the study is based on a 'long-and-thin' dataset, the size of which is limited due to a number of factors. These include the availability of suitable texts, particularly for earlier periods, and the labour-intensive process of manually identifying abbreviations in the corpora. As a result, when it comes to less frequently used lexical words in particular, it is worth considering that examining a larger number of texts from one particular time period could potentially yield different results.

Another limitation of this study is its focus on a single text type for the twenty-first century, despite the existence of a wide range of CMC text types. As mentioned, the choice of IM as the focus of this investigation stems from the existence of previous research into the use of abbreviation within it and the conception of it as a hybrid register that is a good representative of CMC because it exhibits 'a fusion of the full range of variants from the speech community: formal, informal, and highly vernacular' (Tagliamonte & Denis 2008: 3). However, there are a variety of genres and text types within the digital communication sphere and they differ in some fundamental ways. For instance, email can be considered to be relatively formal compared to IM, and is more consistently asynchronous (IM can be both synchronous and asynchronous depending on interaction and language user).

In future studies it would therefore be beneficial to investigate the development of abbreviation practices in a wider range of genres, especially email correspondence, and place it in the context of earlier personal letter writing, although accessing modern email data is challenging due to privacy concerns, and there are very few corpora available as a result. With earlier periods, the problem is the lack of corpora which would have abbreviations encoded. Exploring pre-digital text types known for their brevity could also provide valuable insights. For instance, classified ads in printed newspapers might provide interesting comparative data between printed and digital sources (they are known for their high degree of abbreviation because people had to pay for each character).[8]

---

[8] Urpo Nikanne, personal communication, 24 August 2021.

Furthermore, it is important to acknowledge that text types may affect some of the results. For example, the disproportionately high number of abbreviated names in the ME period is probably caused by text types which contain long lists of witnesses (see section 5.4 above). However, while the higher occurrence of abbreviated names in the ME corpora may be influenced by the kinds of text types included, the study has highlighted consistent changes in how names are abbreviated across periods which are not solely due to the influence of individual text types. The use of brevigraphs in ME and the development towards more efficient, initial-based abbreviations in LModE are consistent trends across periods (see (22)–(24) for developments).

A broad study like this one suggests a whole range of directions for future research. Further investigations could delve into smaller analytic categories within the ones used in the present study. In particular, abbreviations could be studied in more detail by examining their textual functions. Furthermore, while function words are frequently abbreviated throughout, and 'and' is commonly abbreviated in all periods, there does appear to be a shift in relation to the kinds of function words which get abbreviated. In the first two periods, a high number of pronouns such as 'that' are abbreviated, whereas in the later periods, with the rise of elision, auxiliary verbs become more prominent. Additionally, incorporating sociolinguistic variables like gender, class and age, or conducting cross-linguistic comparisons, would be a major priority.

Another interesting approach would be to focus on how each period has its own dynamic and productive abbreviation strategies. For example, the twenty-first-century data show a seemingly spontaneous use of an initialism to abbreviate 'Camp and Furnace' to *C&F*. In the ME period, brevigraphs could be used to abbreviate any lexical item with a notable decrease in the number of brevigraphs and 'other lexical words' in the EModE period. The use of superscript abbreviations in the EModE period could similarly be very productive and still has traces today in the way word-processing software likes to autoformat ordinal numerals into superscript. In contrast to this kind of productivity, it is worth keeping in mind that abbreviations can stay in use for extended periods of time and become fossilised. It would therefore be interesting to systematically compare standardisation and variability in the abbreviations of the same lexeme (see, for example, abbreviations of time in section 5.6).

In conclusion, despite its limitations, this study establishes an important framework. Investigating CMC abbreviation practices using a quantitative, diachronic approach helps to place them in their proper historical context. While many handwritten abbreviations have disappeared, the continued survival of abbreviations like *etc.* or *xmas* demonstrates that they did not all vanish without a trace. Additionally, our study highlights the continuity from LModE to twenty-first-century English, as many of the trends such as the increased use of elision and acronyms are already apparent in the LModE data. Whilst it is important to acknowledge that a study such as this can only hope to scratch the surface, as there are a multitude of factors that need to be studied in more detail, we hope to have laid a foundation and proposed some research questions for future diachronic studies of abbreviation usage.

Authors' addresses:

Department of Literature, Area Studies and European Languages
University of Oslo
PO Box 1003
0315 Oslo
Norway
alpo.honkapohja@gmail.com

Department of English & Creative Arts
Edge Hill University
St Helens Road
Ormskirk L39 4QP
United Kingdom
imogen.marcus@edgehill.ac.uk

Sources

Corpora

Davies, Mark. 2008– *The Corpus of Contemporary American English* (COCA). https://corpus.byu.edu/coca/ (accessed 14 June 2023).

Denison, David. 1994. *A Corpus of Late Modern English Prose*. https://personalpages.manchester.ac.uk/staff/david.denison/lmode_prose.html (accessed 14 June 2023).

Kytö, Merja, Terry Walker & Peter Grund. 2011. *English Witness Depositions 1560–1760: An Electronic Text Edition* (ETED). CD-ROM.

Laing, Margaret. 2013. *A Linguistic Atlas of Early Middle English* (LAEME), 1150–1325. Version 3.2. The University of Edinburgh. www.lel.ed.ac.uk/ihd/laeme2/laeme2.html (accessed 14 June 2023).

Marcus, Imogen & Uschi Maden-Weinberger. 2021. *A Transhistorical Corpus of Written English* (TCWE). www.sketchengine.eu/transhistorical-corpus-of-written-english/ (accessed 14 June 2023).

Nevalainen, Terttu, Helena Raumolin-Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi & Minna Palander-Collin. 1998. *Corpus of Early English Correspondence* (CEEC). www.helsinki.fi/en/researchgroups/variation-contacts-and-change-in-english/research/corpus-of-early-english-correspondence (accessed 14 June 2023).

Stenroos, Merja, Kjetil V. Thengs & Geir Bergstrøm with the assistance of Anastasia Khanukaeva, Delia Schipor & Kenneth Solberg-Harestad. 2020. *A Corpus of Middle English Local Documents* (MELD), version 2017.1. www.uis.no/en/meld-corpus-files (accessed 14 June 2023).

Lexicographical sources

Middle English Dictionary (MED). Ed. Robert E. Lewis *et al*. Ann Arbor: University of Michigan Press, 1952–2001. Online edition in *Middle English Compendium*. Ed. Frances McSparran *et al*. Ann Arbor: University of Michigan Library, 2000–18. http://quod.lib.umich.edu/m/middle-english-dictionary/ (accessed 14 June 2023).

Oxford English Dictionary (OED) online (3rd edition). 2000–. Oxford: Oxford University Press. www.oed.com (last accessed 14 June 2023).

Other primary sources

Benson, Arthur Christopher. 1905. *The Upton Letters.* Project Gutenberg. First Posted 19 February 2002. www.gutenberg.org/files/4615/4615-h/4615-h.htm (accessed 14 June 2023).

Letters to loved ones. Imperial War Museum. www.iwm.org.uk/history/letters-to-loved-ones (accessed 14 June 2023).

References

Álvarez-Gil, Francisco & Marina Bondi. 2021. Introduction to the monographic section: Metadiscourse devices in academic discourse. *Revista Signos: Estudios de Lingüística* 54(106), 518–28.

Amador-Moreno, Carolina. 2019. *Orality in written texts: Using historical corpora to investigate Irish English 1700–1900*. Abingdon: Routledge.

Androutsopoulos, Jannis. 2006. Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics* 10(4), 419–38.

Androutsopoulos, Jannis, 2011. Language change and digital media: A review of conceptions and evidence. In Tore Kristiansen & Nikolas Coupland (eds.), *Standard languages and language standards in a changing Europe*, 145–59. Oslo: Novus Press.

Baron, Naomi. 2004. See you online: Gender issues in college students' use of instant messaging. *Journal of Language and Social Psychology* 23(4), 397–423.

Bassac, Christian. 2004. Principes de morphologie anglaise. *Linguistica*. Pessac: Presses Universitaires de Bordeaux.

Bauer, Laurie. 1993 [1983]. *English word-formation*. Cambridge: Cambridge University Press.

Bergs, Alexander. 2009. Just the same old story? The linguistics of text messaging and its cultural repercussions. In Charley Rowe & Eva Lia Wyss (eds.), *Language and new media: Linguistic, cultural, and technological evolutions*, 55–73. Cresskill, NJ: Hampton.

Biber, Douglas & Edward Finegan. 1989. Drift and the evolution of English style: A history of three genres. *Language* 65(3), 487–517.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 2000. *Longman grammar of spoken and written English*. London: Longman.

Bieswanger, Markus. 2013. Micro-linguistic structural features of computer-mediated communication. In Susan Herring, Dieter Stein & Tuija Virtanen (eds.), *Pragmatics of computer-mediated communication*, 463–85. Berlin: De Gruyter Mouton.

Bieswanger, Markus. 2016. Electronically-mediated Englishes: Synchronicity revisited. In Lauren Squires (ed.), *English in computer-mediated communication: Variation, representation, and change*, 281–300. Berlin: De Gruyter Mouton.

Cannon, Garland. 1989. Abbreviations and acronyms in English word-formation. *American Speech* 64(2), 99–127.

Cappelli, Adriano. 1899. *Lexicon abbreviaturarum dizionario di abbreviature latine ed italiane*. Milan: Hoepli.

Crystal, David. 2001. *Language and the Internet*. Cambridge: Cambridge University Press.

Culpeper, Jonathan & Merja Kytö. 2010. *Early Modern English dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.

Danet, Brenda. 2001. *Cyberpl@y: Communicating online*. Abingdon: Routledge.

Drouin, Michelle & Claire Davis. 2009. R u txting? Is the use of text speak hurting your literacy? *Journal of Literacy Research* 41(1), 46–67.

Eberl, Martin. 2020. Double trouble. In Sofia Rüdiger & Daria Dayter (eds.), *Corpus approaches to social media: Studies in corpus linguistics*, 131–46. Amsterdam: John Benjamins.

Elspaß, Stephan. 2002. Alter Wein und neue Schläuche? Briefe der Wende zum 20. Jahrhundert und Texte der neuen Medien – ein Vergleich. *Osnabrücker Beiträge zur Sprachtheorie* 64, 7–32.

Fitzmaurice, Susan. 2002. *The familiar letter in Early Modern English: A pragmatic approach*. Amsterdam: John Benjamins.

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 Michel Banniard (ed.), *Langages et peuples d'Europe: Cristallisation des identités romanes et germaniques (VIIe–Xie siècle)*, 79–110. Toulouse: CNRS Université de Toulouse-Le Mirail.

Hector, Leonard. 1958. *The handwriting of English documents*. Ilkley: Scolar Press.

Herring, Susan C. 2005. Computer-mediated communication on the internet. *Annual Review of Information Science and Technology* 36(1), https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/aris.1440360104 (accessed 14 June 2023).

Heuman, Anna. 2021. Acceptable language online: Negotiating the impact of digital affordances. In Marjut Johansson, Sanna-Kaisa Tanskanen & Jan Chovanec (eds.), *Analyzing digital discourses: Between convergence and controversy*, 163–87. London: Springer Nature.

Honkapohja, Alpo. 2013. Manuscript abbreviations in Latin and English: History, typologies and how to tackle them in encoding. *Studies in Variation, Contacts and Change in English* 14, https://varieng.helsinki.fi/series/volumes/14/honkapohja/ (accessed 14 June 2023).

Honkapohja, Alpo. 2021. Digital approaches to manuscript abbreviations: Where are we at the beginning of the 2020s? *Digital Medievalist* 14, https://journal.digitalmedievalist.org/article/id/7025/ (accessed 14 June 2023).

Honkapohja, Alpo & Aino Liira. 2020. Abbreviations and standardisation in the *Polychronicon*: Latin to English and manuscript to print. In Laura Wright (ed.), *The multilingual origins of standard English*, 269–316. Berlin: De Gruyter Mouton.

Jamet, Denis. 2009. A morphological approach to clipping in English: Can the study of clipping be formalized? *Lexis: Journal in English Lexicology* (HS 1), https://journals.openedition.org/lexis/884 (accessed 11 March 2023).

Jucker, Andreas H. 2020. *Politeness in the history of English: From the Middle Ages to the present day*. Cambridge: Cambridge University Press.

Katamba, Francis. 2005. *English words. Structure, history, usage*, 2nd edn. Abingdon: Routledge.

Ker, Neil R. 1960. *English manuscripts in the century after the Norman Conquest*. Oxford: Clarendon.

Kestemont, Mike. 2015. A computational analysis of the scribal profiles in two of the oldest manuscripts of Hadewijch's letters. *Scriptorium* 69, 159–75.

Kjeldsen, Alex Speed. 2013. *Filologiske studier i kongesagahåndskriftet Morkinskinna*. Copenhagen: Museum Tusculanum Press.

Klymenko, Olga. 2019. Twitterverse: The birth of new words. *Proceedings of the Linguistic Society of America* 4(1), 1–11.

Lang, Tom. 2019. The long and the short of abbreviations. *European Science Editing* 45(1), 152–60.

Lowe, Kathryn. 2006. Palaeography. In Keith Brown (ed.), *Encyclopedia of language and linguistics*, 2nd edn, 134–41. London: Elsevier.

Lyddy, Fiona, Francesca Farina, James Hanney, Lynn Farrell & Niamh Kelly O'Neill. 2014. An analysis of language in university students' text messages. *Journal of Computer-Mediated Communication,* 19(3), 546–61.

Marcus, Imogen. 2017. *The linguistics of spoken communication in Early Modern English writing: Exploring Bess of Hardwick's manuscript letters*. Cham: Springer.

McCulloch, Gretchen. 2019. *Because internet: Understanding how language is changing*. London: Vintage.

Metcalf, Allan. 2010. *OK: The improbable story of America's greatest word*. Oxford: Oxford University Press.

Moore, Colette. 2006. The use of *videlicet* in Early Modern slander depositions: A case of genre-specific grammaticalization. *Journal of Historical Pragmatics* 7(2), 245–63.

Nevala, Minna. 2004. Accessing politeness axes: Forms of address and terms of reference in early English correspondence. *Journal of Pragmatics* 36, 2125–60.

Palander-Collin, Minna. 2010. Correspondence. In Andreas H. Jucker & Irma Taavitsainen (eds.), *Historical pragmatics*, 651–78. Berlin: De Gruyter Mouton.

Petti, Anthony. 1977. *English literary hands from Chaucer to Dryden*. London: Arnold.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Quan-Haase, Anabel. 2009. Text-based conversations over instant messaging: Linguistic changes and young people's sociability. In Charley Rowe & Eva Lia Wyss (eds.), *Language and new media: Linguistic, cultural and technological evolutions*, 33–54. Cresskill, NJ: Hampton.

Rodríguez, Félix González & Garland Cannon. 1994. Remarks on the origin and evolution of abbreviations and acronyms. In Francisco Moreno Fernández, Miguel Fuster & Juan José Calvo (eds.), *English historical linguistics 1992*, 261–72. Amsterdam: John Benjamins.

Rutten, Gijsbert & Marijke J. van der Wal. 2014. *Letters as loot: A sociolinguistic approach to seventeenth- and eighteenth-century Dutch*, vol. 2. Amsterdam: John Benjamins.

Schneebeli, Célia. 2020. Where *lol* is: Function and position of *lol* used as a discourse marker in YouTube comments. *Discours: Revue de linguistique, psycholinguistique et informatique* 27, https://journals.openedition.org/discours/10900 (accessed 14 June 2023).

Shortis, Tim. 2016. Texting and other messaging: Written system in digitally mediated vernaculars. In Vivian Cook & Des Ryan (eds.), *The Routledge handbook of the English writing system*, 507–36. Abingdon: Routledge.

Stam, Nike. 2017. A typology of code-switching in the commentary to the *Félire Óengusso*. PhD dissertation, University of Utrecht.

Stenroos, Merja, Geir Bergstrøm & Kjetil V. Thengs. 2020. The categorization of Middle English documents: Interactions of function, form and language. In Merja Stenroos & Kjetil V. Thengs (eds.), *Records of real people: Linguistic variation in Middle English local documents*, 37–68. Amsterdam: John Benjamins.

Stockwell, Robert & Donka Minkova. 2003. *English words: History and structure*. Cambridge: Cambridge University Press.

Tagliamonte, Sali A. & Derek Denis. 2008. Linguistic ruin? LOL! Instant messaging and teen language. *American Speech* 83(1), 3–34.

Traube, Ludwig. 1907. *Nomina Sacra: Versuch einer Geschichte der christlichen Kürzung*, vol. 104. Munich: C. H. Beck'sche Verlagsbuchhandlung.

Williams, Joseph M. 1981. *Style: Ten lessons in clarity and grace*. Glenview, IL: Foresman.

Zelenkauskaite, Asta. 2017. Abbreviate and insert? Message length, addressee and non-standard writing in Italian mobile texting and Facebook. *Writing Systems Research* 9(2), 123–36.