UNIVERSITY OF MURCIA

International Journal of English Studies

IJES

http://revistas.um.es/ijes

Early Modern Medicine in Manuscript and Print: A Triangulation Approach to Analysing Spelling Standardisation

JUKKA TYRKKÖ[*]

Linnaeus University (Sweden)

Received: 06/04/2020. Accepted: 15/06/2020.

ABSTRACT

The standardisation process of English spelling largely came to its conclusion during the Early Modern period. While the progress of standardisation has been studied in both printed and manuscript texts, few studies have looked at these processes side by side, especially focusing on the same genre of writing and by using corpora that are sufficiently large for quantitative comparison. Using two Early Modern medical corpora, one based on manuscripts and the other on printed sources, this paper compares the trajectories of spelling standardisation in the two textual domains and shows that while spelling standardisation progressed in an almost linear fashion in printed texts, the manuscripts reveal a much more varied and shallow cline toward standardisation.

KEYWORDS: Printing; Manuscript; Spelling; Standardisation; Early Modern; Statistics.

1. INTRODUCTION

Over the last twenty years, the improved diversity of historical corpora has fundamentally changed the way we study language history. The use of digital methods in the exploration of these diverse datasets has allowed us to pinpoint crucial turning points and identify trajectories of change in ways that were never possible before. Likewise, we can use pattern-driven methods to test previously formulated theories and models on new datasets to gain a fuller picture of the universality of linguistic phenomena and to compare different datasets in order

[*]***Address for correspondence***: Department of Languages, Faculty of Arts and Humanities, Linnaeus University; e-mail: jukka.tyrkko@lnu.se.

Print ISSN: 1578–7044; Online ISSN: 1989–6131
doi: 10.6018/ijes.421911

to identify key differences. The topic at hand in this paper concerns spelling standardisation in Early Modern medical writing with special reference to differences between the progress of standardisation in printed and hand-written medical books (for a good overview of spelling in historical scientific writing, see Taavitsainen [2000]).

This paper explores three complementary approaches to quantitative analysis of spelling standardisation, namely *semi-automatic spelling normalisation*, *lexical diversity* and *n-gram-based frequency profiling*. These methods will be used to analyse and compare two medical corpora of the Early Modern period, one representing printed medical writing (*Early Modern English Medical Texts* [Taavitsainen et al., 2010], or EMEMT) and the other representing hand-written medical books from the same time period (*The Málaga Corpus of Early Modern English Scientific Prose* [Calle-Martín et al., 2016–], or MCEMESP); see section 3 for descriptions of the corpora. As far as the author is aware, this study is the first direct comparison of spelling in contemporary scientific manuscripts and print publications; for contrastive studies of manuscript and print, see e.g. Boffey (2012), Marotti (1995) and Sönmez (1993). The fact that the corpora represent the same genre is important, particularly from the perspective of both semi-automated normalisation methods and quantitative analysis, because Early Modern genres varied greatly in the amount of multilingual content, and genres like medicine contain sometimes considerable amounts of text in Latin and Greek (see Pahta, 2011).

In addition to the issue of genre, another important consideration concerns the accuracy of the corpus transcription, perhaps in particular when it comes to manuscript corpora. As Sairio et al. (2018: 80) note, the majority of what we know about early spelling comes from printed and publicly distributed sources, largely due to the fact that large-scale datasets of diplomatically transcribed private writings, such as letters and other autograph texts, have been relatively scarce and limited in scope. Although such texts have been compiled and published in reasonably large numbers, it was until recently very common for editors to modernise the texts, often silently. According to Sairio et al. (2018: 80), "… many editions claim to retain original spellings, in practice editors always normalize texts at the graph level, and some practices of silent editorial normalization have become so conventionalized that they may not even be mentioned in the editorial principles". By contrast, the transcription in the manuscript corpus used here is of exceptional quality, which facilitates the contrastive approach taken in this study.

Although the findings will be discussed against a background of philological and book historical knowledge about factors that have, or have been claimed to have, affected Early Modern spelling, the focus of the paper is decidedly quantitative and exploratory, and the historical details of spelling standardisation are mostly addressed through reference to earlier scholarship. The main objective of the analysis is to explore the different trajectories of spelling standardisation in the two text types, one directly affected by the *community of practice* of London printers while the other was not (Tyrkkö, 2013). Data visualisation methods will be used throughout to make the quantitative findings easier to interpret.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

2. BACKGROUND

In linguistics, the term *standard* is typically used in reference to standard varieties, that is, varieties of languages, and features thereof, that are commonly used in public contexts, institutionally supported, and are considered unmarked. The opposite of a standard variety is a non-standard variety, a variety that may be regionally dominant, but is considered inappropriate in public (and certainly official) contexts, is not taught (and often proscribed against) in schools, and is marked, typically as being a low prestige variety. In light of these common definitions, the concept of *non-standard spelling* can be understood both synchronically and diachronically, and it may be useful to make a distinction between the two. Synchronically, non-standard spelling means that there is an established way of spelling a word, and that in a particular instance a lexical concept is given a form that (intentionally) differs from the standard spelling. For example, in so-called Internet English, the word "cause", a shortened informal rendering of "because", may be further clipped and given as "cuz". By contrast, diachronic non-standard spelling, which could and perhaps should be called *pre-standard spelling*, refers to spelling practices that differ from our current practice. Notably, pre-standard spelling was actually often, though not always, standard spelling for the time. The use of "u" where we would today employ "v" –for example, "euery" for "every"– was perfectly standard through all of the sixteenth century and into the seventeenth; see, e.g. Rutkowska (2013, 2016 and 2017).

Space does not permit even an abbreviated historical overview of the timeline and major events of English spelling standardisation in this paper. Suffice it to say that the topic has been discussed extensively, along with other aspects of the standardisation of English, by numerous excellent scholars, e.g. Moessner (2017), Nevalainen (2012), Nevalainen and Tieken-Boon van Ostade (2006), Scragg (1974), Stenroos and Smith (2016), and Tieken-Boon van Ostade (1998). However, although the impact of the printing revolution to spelling standardisation is well acknowledged, and has been discussed frequently, a few words will be devoted to that particular aspect because it pertains directly to the task of comparing spelling standardisation in printed and manuscript sources.

Much has been made of the role of the early printers in shaping the direction of English spelling, and it is evidently clear that this was indeed the case (see e.g. Tyrkkö, 2013). Not only did the printing process accelerate regularisation, but the increase in exposure to texts no doubt affected contemporary readers over time when it came to spelling; the overall impact of printing has been discussed extensively in literature, especially since Eisenstein (1979). However, the extent to which spelling standardisation was planned by individual printing houses, on the one hand, or coordinated among them, on the other, remains much less clear. In order to assess this, we need to consider the reality of book printing in London during the Early Modern period, and while the actual circumstances and practices of book printing may fall

Print ISSN: 1578–7044; Online ISSN: 1989–6131

more into the purview of book history proper, there are several aspects that are worth mentioning here.

Firstly, it is important to remind ourselves that the printing industry was a highly-localised affair in England throughout much of the Early Modern period. Nearly all important printing took place in London and even there much of it within a relatively concentrated area. Most printing shops were located in the vicinity of St. Paul's Churchyard, and the number of establishments was small enough to allow everyone in the industry to know each other; I have argued elsewhere that the community-of-practice perspective is a valuable one to take when it comes to Early Modern printing and thus to spelling standardisation (Tyrkkö, 2013). The printers formed a tight-knit professional community where new innovations, or deviations from current standards, were immediately noticed. itIndeed, given that London printing-houses typically employed anywhere from five to fifteen people (Bland, 2010: 107) and only some of these had a direct primary influence on spelling –namely the master printer, the compositors and the correctors–, which means that radically new ways of doing things could perhaps be associated with specific individuals. However, in reality book printing was a much more complicated affair, and most attempts at so-called compositor identification have met with relatively limited success; see e.g. McKerrow (1967: 23). As Bland (2010: 108–112) points out, not only could several teams of compositors and typesetters work on the same book at the same time, but printing houses would also join forces in a variety of different ways, and this is only occasionally indicated in the imprint of the book. It is therefore almost impossible to identify specific printing houses, let alone individuals, who had a particular effect on standardisation.

Secondly, we need to consider why printers would have been invested in the standardisation of spelling. Howard-Hill (2006: 18–19) has argued that the emergence of standardisation was essentially an economically driven process, while Rutkowska (2013) notes that "the systematisation of orthographic principles seems to have contributed to the increased efficiency and effectiveness of printers' work". Regularly-spelled text is simply quicker and more efficient to set than following the idiosyncratic spellings of individual authors, and for a printing house nothing was more important than getting books out and in the hands of customers. In addition to proofreading text, one of the duties of the corrector was often to aid the author by silently correcting their language as well, which included spelling (see Grafton, 2012).

It is wrong, however, to say that printers had full control over the manuscript the moment it was handed over by the author, because it is clear that printers did take authors' own spellings into consideration, particularly in the case of more prominent authors (McKerrow, 1967: 247). While contemporary printers' manuals like Moxon's famous *Mechanick Exercises* (1683) go into detail about revisions and changes that may be made to the text, they typically concern minor visual changes to aid the reader and make little reference to spelling and none at all that would concern standardisation; see Maruca (2003: 330–338). Nor, it goes without saying, were

Print ISSN: 1578–7044; Online ISSN: 1989–6131

printers immune to the contemporary discourse surrounding spelling, which engaged both authors like Swift, Addison, Steele and Defoe (see e.g. Neumann, 1944) and scholars like the spelling reformer John Hart and the conservative Richard Mulcaster (see Doval, 1996). This is not to claim that London printers would have taken either the authors or the spelling reformers concerns to heart, but it is clear that particularly the more notable printing houses took great pride in their product, and as the decades progressed, the external pressure to conform to the developing standard was no doubt considerable.

When it comes to hand-written texts, it is of course self-evident that greater variation follows from idiographic practice (see e.g. Kestemont, 2015). Not only could individual authors of informal texts spell words in an idiosyncratic fashion (see Osselton, 1984), if they wished to do so, but they were also free to make use of letters and abbreviations which simply could not be produced in print due to the lack of necessary typesets –after all, most London printers ordered their type from the continent and were therefore largely limited to whatever type was available. According to Honkapohja and Liira (2020), in Middle English manuscripts the proportion of abbreviated words could reach more than 50% in specific sections, and although the use of abbreviations declined during the Early Modern period, they continued to be used throughout.

3. MEDICAL WRITING AND THE TWO MEDICAL CORPORA

In England, medicine was the first field of scientific writing that saw a turn to the vernacular starting in the Medieval period and gradually overtaking Latin during the Early Modern period. Although Latin remained the language of the relatively small community of university-trained physicians, the vast majority of medical practitioners who were trained through apprenticeship –apothecaries, surgeons, midwives, bonesetters and others– either knew no Latin at all or certainly too little to write anything in the language. This fact makes medical writing a particularly attractive genre of writing to study from the perspective of language change, including spelling (see Taavitsainen, 2000).

This study makes use of two corpora of Early Modern medical writing: *Early Modern English Medical Texts* (EMEMT) and *The Málaga Corpus of Early Modern English Scientific Prose* (MCEMESP). Covering the same time period and similar groups of authors, the two corpora differ primarily in the mode of communication. The texts in EMEMT represent printed medical writing while those in MCEMESP cover medical manuscripts. All the texts are authentic medical texts written by medical practitioners. By being able to study texts of the same genre in print and manuscript allows us to compare like with like, instead of comparing spelling in printed books or periodicals with private correspondence. So far, studies covering medical writing in both modes are rare (see e.g. Pahta et al., 2011), but the introduction of corpora like MCEMESP facilitates these new investigations. Details of individual texts in the

Print ISSN: 1578–7044; Online ISSN: 1989–6131

two corpora can be found in the respective manuals of the corpora (Taavitsainen et al., 2011 and https://modernmss.uma.es/Corpus).

3.1. *Early Modern English Medical Texts* (EMEMT)

The EMEMT corpus covers a broad range of different types of printed medical texts from the period 1500 to 1700.[1] The corpus includes 175 extracts of c. 10,000 words from medical books as well as a selection of medical articles from the *Philosophical Transactions of the Royal Society*, founded in 1660. The texts were selected to represent Early Modern vernacular medical writing across different subgenres and topics of contemporary medicine. Although some publishers were known to specialise in learned topics more than others, and some were more likely to publish books that needed high-quality engravings, medical books were printed by a relatively wide range of printing houses. This is naturally important for the present study, because if several extracts came from a single printing house, the standards and routines of that printing house could skew the diachronic trends that are of interest here.

The size of the full corpus is 1,812,932 tokens, of which 1,104,343 come from the 124 medical books included in the present study. These texts come from three of the six major 'categories' used in EMEMT,[2] namely *specialised treatises*, *remedy books* and *surgical books*. General treatises, recipe collections and articles published in the *Philosophical Transactions* were excluded. General treatises were excluded because they comprise both learned and recipe collections as their lexical composition is quite different from vernacular English, and the *Philosophical Transactions* because they represent a new genre, scientific periodicals, that emerged during the 1660s and therefore does not extend over the full time.

The texts in EMEMT were keyed-in by the project members from facsimile copies of the books. Each corpus file was proofread twice by other project members, and checked against an original copy held by a scholarly library, in most cases the British Library or Wellcome Library. The transcriptions retained the original spellings and lineations, but some items were standardised as part of the editing process (for details, see Tyrkkö et al., 2010). These editorial interventions included converting the long <ſ> into standard <s>, etc. Brevigraphs were annotated into EMEMT files using custom markup. In print, the use of brevigraphs was relatively limited due to contemporary printing technology, and the vast majority of brevigraphs involve superscript characters indicated in the corpus with the annotations "=e=" and "=t=". The most typical example is the group <the>, or definite article, which occasionally appears in earlier printed texts as the character <y> with a raised minuscule <e> directly above. The <y> represented the character thorn (<þ>), which had largely been replaced by the digraph and which printers rarely had access to in their type fonts (see Tyrkkö, 2013: 159).[3] Macrons also appear with some regularity in early printed texts.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

A spelling-standardised version of the corpus was prepared using VARD (see section 4.1) and released as a parallel version at the time the corpus was published (see Lehto et al., 2010). This version was used in the study, as discussed in section 4.1.

3.2. *The Málaga Corpus of Early Modern English Scientific Prose* (MCEMESP)

The MCEMESP corpus comprises 13 diplomatic transcriptions of hand-written medical books.[4] The manuscripts are housed at the Hunterian Collection at Glasgow University Library, the Wellcome Collection at the Wellcome Library in London and the Rylands Collection at the University of Manchester Library. The full corpus comprises 498,718 tokens. The texts represent the manuscript tradition of learned medical writing, which was still fully vibrant during the Early Modern period despite the increasing availability of printing.

The transcriptions were produced manually and they are of very high quality. The editorial intervention was minimal, and the compilers retained spellings, capitalisations, word divisions and punctuations of the original texts. Importantly for the present study, unlike some other corpus projects, the compilers did not silently standardise characters such as <u> for <v> and <i> for <j>, which preserves access to the authentic orthography of the manuscripts. For more information, see e.g. Romero-Barranco (2017).

At the time the present study was finalised, spelling standardisation using VARD had been carried out on 10 of the 13 files. Although it would have been possible to produce the missing files for the purposes of this study, I opted to use only those files that are made officially available by the corpus compilers in order to ensure verifiability. The corpus provides general datings for the manuscripts, which were converted into approximate pseudo-precise datings for the purposes of the visual data exploration: for example, "sixteenth century" would have been interpreted as the middle of the century and thus represented as 1550. More precise datings, which are of course usually impossible with manuscripts, would improve the accuracy of the trajectory analyses.

4. METHODS

The orthographic differences between the medical books and the manuscripts will be examined by means of triangulation using a three-complementary method which relies on computational and quantitative analysis, as well as a variety of data visualisation techniques. These methods and their results are presented here not as a replacement, but as a potential complement to tried-and-tested qualitative approaches. In particular, the argument is made that it is important to pay attention to textual level variation when we study linguistic and other related processes of change, rather than binning data from diverse individual texts into periods, genres or other pre-determined groupings.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

4.1. Automatic spelling standardisation

Automatic spelling standardisation is a method of computational linguistics, which applies a series of algorithms to texts in order to recognise non-standard spellings and then regularises them, typically using Present-day spellings as the target. The resulting texts are effectively similar to traditional modernised editions of historical books. Spelling standardisation has been found to be particularly useful in historical corpus linguistics, especially in corpus-driven studies that benefit from differently spelled forms of the same words being processed together: for example, collocation extraction, keyword analysis and n-gram analysis all require standardised spellings.

Perhaps the most well-known spelling standardisation tool is the Variant Detector, or VARD (Baron, 2011); for recent developments in the field, see Hämäläinen et al. (2018) and Pettersson et al. (2013). Developed by Alistair Baron, VARD is widely used in historical corpus linguistics for producing spelling-standardised versions of corpora (e.g. Lehto et al., 2010).[5] VARD comes with built-in rules for standardisation, but the tool can be manually trained for specific time periods and text types, as well as to customise the standardisation for specific purposes. For example, although VARD can normalise the second-person singular "thou" as "you", most compilers elect to retain the former forms. In practice, VARD provides as estimate of reliability as well as suggestions of possible standardised forms, and learns the correct forms supplied by the user.

It is important to reiterate here that the objective of the present study is not to evaluate the accuracy or reliability of automatic spelling standardisation tools, but rather to explore methods that may be used to evaluate corpora that have been annotated with such tools. In the present case, VARD-standardised versions are available of both MCEMESP and EMEMT, and thus the choice of tool was predetermined by the circumstances. Both spelling standardised versions were prepared by the respective teams of compilers; more information on the VARDing of EMEMT is available in Lehto et al. (2010). The standardised versions of the two corpora were used as provided with no further post-processing by the present author. The method used in the present study was first discussed in Tyrkkö (2013). In short, VARD's output of markup is used as the basis of the analysis by quantifying the proportional volume of non-standard spellings for each text, and then examining the diachronic distributions of these proportions.[6] Given that nearly ten years passed between the standardisation of the two corpora and two different versions of the software were used –EMEMT with VARD, MCEMESP with VARD2– the two versions differ somewhat in terms of VARD markup; see examples 1 and 2.

> (1) MInding to <replaced orig="answeare">answer</replaced> the little <replaced orig="booke">book</replaced> of Aubertus, concerning the <replaced orig="originall">original</replaced> of <replaced orig="mettals">metals</replaced>, and their causes, (EMEMT, DuChesne 1591, f.1r)

Print ISSN: 1578–7044; Online ISSN: 1989–6131

(2) The Secrets and Experiments of <notvariant>Mathewe</notvariant> <notvariant>Lucatell</notvariant><normalised orig="þe auto="false">the</normalised> Mister <normalised orig="Chirurgion" auto="false">chirurgeon</normalised> of Venice. (MCEMESP, Ms Hunter 43, p. 1)

4.2. Lexical diversity and lexical similarity analysis

The second question relevant to the study of spelling standardisation is lexical diversity. If we count each instance of a unique token (or word form) as a lexical item without collapsing all supposed variants of the same lexical item under the corresponding standardised form, does lexical diversity in texts decline as standardisation proceeds? Or, to put it another way, are older texts more diverse when it comes to unique tokens on the page?

A number of different descriptive statistics have been proposed for quantifying lexical richness. Whilst the general principle is the same as in the commonly known *type-token ratio* (TTR), namely to assess the amount of different words or word forms occurring in a text, different methods have been proposed for dealing with varying lengths of texts, which naturally affect the measure. In the present study, we will use VoC-D, or Vocabulary Density, introduced by McKee et al. (2000), which is very robust when it comes to text length. Unlike the more familiar type-token ratio or variations thereof, the statistic D is calculated by finding the best fit for two curves, one based on a random sampling of words and the other on the TTR ratio. The full details of the method are available in McKee et al. (2000).

As the term implies, lexical similarity is a measure of how similar two word forms are. Different methods of similarity measuring are used in predictive typing, autocorrection of spelling, and many other applications, but they can also be used for quantifying variation. The similarity or dissimilarity of two strings of letters can be conceptualised as an *edit distance*, or how many changes, insertions and deletions would have to be made to string A in order to turn it into string B? For example, turning *cat* into *bat* requires one change (*c* into *b*), while turning *cat* into *battle* requires four (changing *c* into *b* as well as the insertion of *t*, *l* and *e* to the end of the word form). Edit distance can be quantified using metrics such as *Levenshtein distance*, which is calculated as the sum of the number of insertions, deletions and substitutions required to turn string A into string B (see Pettersson et al., 2013). In the present case, Levenshtein distance was used to quantify how different the non-standard variants are from their standard variants. The non-standard and corresponding standard forms were derived from VARD and the edit distances were calculated for each pair, with Roman numerals excluded from the inventory. To account for the different word lengths, the Levenshtein scores were standardised and the mean Levenshtein scores were calculated for each text based on all the non-standard words. The higher the score, the fewer edits were needed on average to change the non-standard form into the standard form.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

4.3. Character n-gram extraction

The term *n-gram* is widely used in reference to sequences of words in a variety of theoretical and methodological contexts, typically with an emphasis on sequences that are frequently occurring or repetitive within a given text or in a corpus (see e.g. Brown [2013] and Heikkilä & Roos [2018]). By the same token, a *character n-gram* is a sequence of characters occurring in text: for example, the word "character" could be understood to comprise the three-grams "cha", "har", "ara" and so on. Depending on the objectives of the analysis, the optimal length of the n-grams extracted can differ, the principles of extracting the n-grams can be adjusted, and the ensuing calculations can be modified in a variety of ways; for example, the exact positions of n-grams within words may be noted.

Languages differ in their inventories of character n-grams, both in terms of what specific n-grams are found and their frequencies. This fact has given rise to a variety of language identification algorithms based on character n-grams. Even closely related languages and dialects of the same languages typically differ to a significant degree at the character n-gram level, and similar methods can be used to examine historical variants of the same language or, as will be done in the present study, texts derived from different media.

The trigram analysis was carried out by using a simple Livecode script that collected every unique sequence of three characters in each text and counted their frequencies per text. The retrieval method iterated over all words of three words or longer, collecting the word-initial trigram and the word-final trigram.

5. FINDINGS

5.1. Spelling standardisation

The first step of the quantitative analysis was to examine the frequencies of non-standard items on the timeline on a text-by-text basis. In order to do this, the number of items identified by VARD as non-standard were normalised per 1,000 words on the vertical axis, with the horizontal axis representing the timeline in years. Each marker represents one text, with the circle representing the manuscript and the plus sign (+) representing the printed texts. Note that many of the manuscripts are dated only approximately using a range, and in the plot they are placed in the middle of the range; for example, "seventeenth century" becomes 1650. The linear lines of best fit and confidence bounds are plotted to show the general trends. A visual representation of the data such as this is arguably much easier to interpret, it allows us to get a 'bird's-eye view' of the distribution of individual items and the trends they follow, and also to identify possible outliers (Figure 1).

Print ISSN: 1578–7044; Online ISSN: 1989–6131

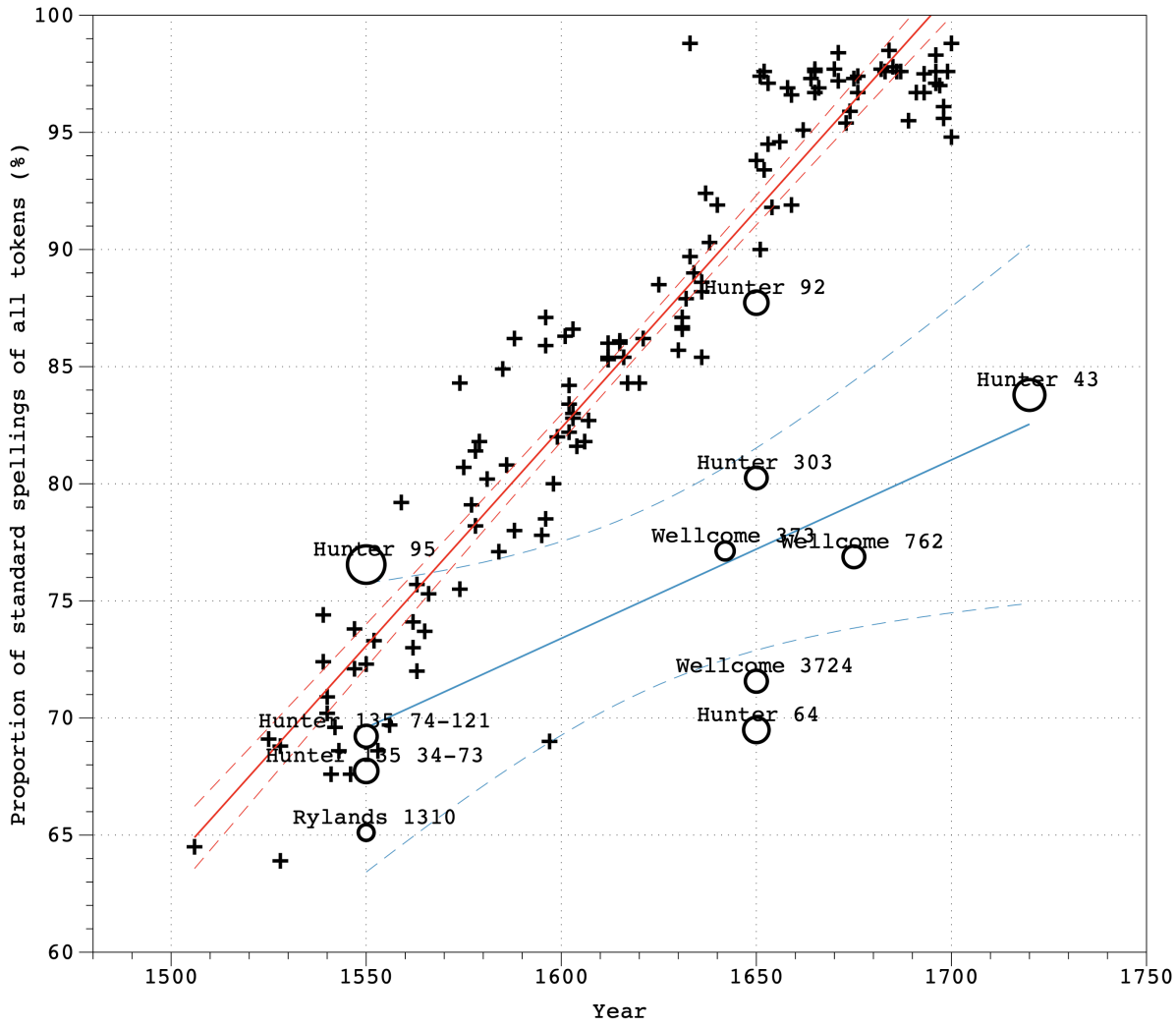

Figure 1. Trajectories of standardisation in EMEMT and MCEMESP.

As Figure 1 shows, the gradual standardisation of spelling is evident in both the manuscript and print datasets; for discussion of differences between the different subgenres of medical printing, see Tyrkkö (2013: 162–165). The diachronic trend of the proportion of non-standard spellings in the printed texts agrees notably well with previous quantitative findings based on a large corpus of texts from the *Early English Books Online* (EEBO) database (see Baron, 2011: 54).

Although the starting point for both text types in the middle of the sixteenth century appears to be substantially the same, c. 65–70% standard spellings, the process of standardisation is shown to proceed much more gradually in the manuscript corpus and also to exhibit much greater variation.[7] Even though the relatively small number of manuscripts in the Málaga corpus makes the trend less reliable –as indicated by the width of the 95% confidence bound– it is clear that there is a substantial and significant difference between the two datasets. While the manuscript texts continue to show relatively high frequencies of non-standard

Print ISSN: 1578–7044; Online ISSN: 1989–6131

spellings well into the eighteenth century, the printed texts achieve what we would consider modern standards by the late seventeenth century. This observation is more or less in line with Scragg (1974: 68), who argued that standardised spelling was achieved by the middle of the century.

The uniformity of the trend in printed texts is quite noteworthy, and it serves as evidence of the extent to which early modern London printers formed a community of practice. Although the Stationer's Company did not prescribe formal regulations about spelling, the printers observed each other's work and made similar changes themselves. As Bland (2010: 119) notes, the conventions must have seemed natural and instinctive to members of the printing community, who were not only intimately familiar with each other's work, but also knew each other socially. If we compare the trend we see in printed books with the findings from the manuscript corpus, we see that there appears to be much more diversity between the data point in the latter, particularly during the seventeenth century. Focusing on the middle of the seventeenth century, we can see that the mean frequency of non-standard spellings is around 225 for the manuscripts and around 50 for the printed texts. The magnitude of the difference, nearly 450%, is striking.

However, in order not to oversimplify matters, it is necessary to look closer at the individual texts. Although the overall difference between printed and manuscript texts is clear, it is noteworthy that some scribes appear to be following the emerging spelling standards fairly closely, while others do not. In particular, Hunter 92, *Treatise on the Anatomy of the Eye*, by the anatomist and surgeon John Browne, shows a rate of 87.7% of standard spellings. This figure is only marginally lower than what we see in contemporary printed books. Browne was infamous as one of the first noted scientific plagiarists, with most of his works being pirated from previously published works. It may have been a manuscript for a print edition, but none appears to have been produced. Given the background, one possible explanation for the surprisingly standard spelling in Hunter 92 is that it might be based on printed works, either fully or in part. The latter possibility would be made more likely if there was significant variation within the manuscript. To test this, Hunter 92 was experimentally split into 10 parts and the normalised frequency of non-standard spellings was calculated in each part; the result showed a uniform distribution of non-standard spellings within the manuscript.

By contrast, the five nearly contemporaneous later manuscripts, Wellcome 373, Wellcome 762, Wellcome 3724, Hunter 303 and Hunter 64, feature more than two times as many non-standard spellings. Hunter 64 and Wellcome 3724, in particular, feature only 70% and 74% standard spellings, respectively, which is reminiscent of manuscripts and printed texts from a century earlier. The youngest manuscript in the corpus, Hunter 43, which is dated between 1679 and 1755, shows 83.8% of standard spellings. Here, too, we witness a rate of non-standard spellings that would have been normal for printed texts more than a hundred years earlier. Compared to contemporary printed texts, the non-standard spellings are nearly ten times more frequent.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

The ratio of spelling variants per standard form informs us about the fixedness of non-standard forms, and therefore about the consistency of spelling regardless of whether the chosen spelling ended up becoming the standard variant we use today. Figure 2 gives an impression of the overall distribution as a treemap. Individual texts are represented as clusters of blocks, the blocks represent individual lexical items, and the sizes of the blocks indicate the relative numbers of different non-standard spellings per lexical item; items featuring only standard spelling were excluded. The later texts with very few non-standard spelling variants disappear into the lower right-hand corner of the image. The colours indicate years, with orange for earlier years and reds for later years, ranging from the beginning of the sixteenth century to the end of the seventeenth century.

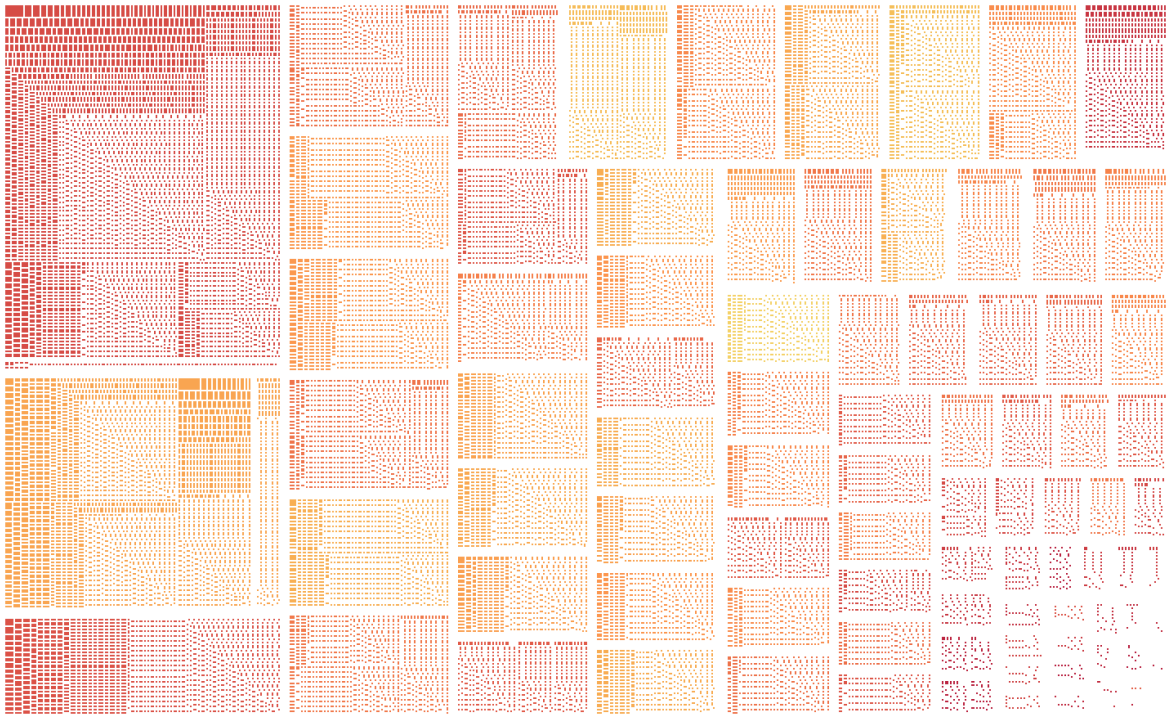

Figure 2. Treemap of spelling variants per text (both corpora).

In Figure 3, a ratio of 1 would indicate that there was one non-standard spelling per corresponding lexical item, while a ratio greater than 1 indicates that each lexical item that was spelled in a non-standard way featured more than one non-standard spelling. As can be seen, the ratio declines over time in both the manuscripts and printed texts, but it is consistently higher in the manuscripts throughout the timeline. The data were first calculated by text and then binned into periods of 20 years.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

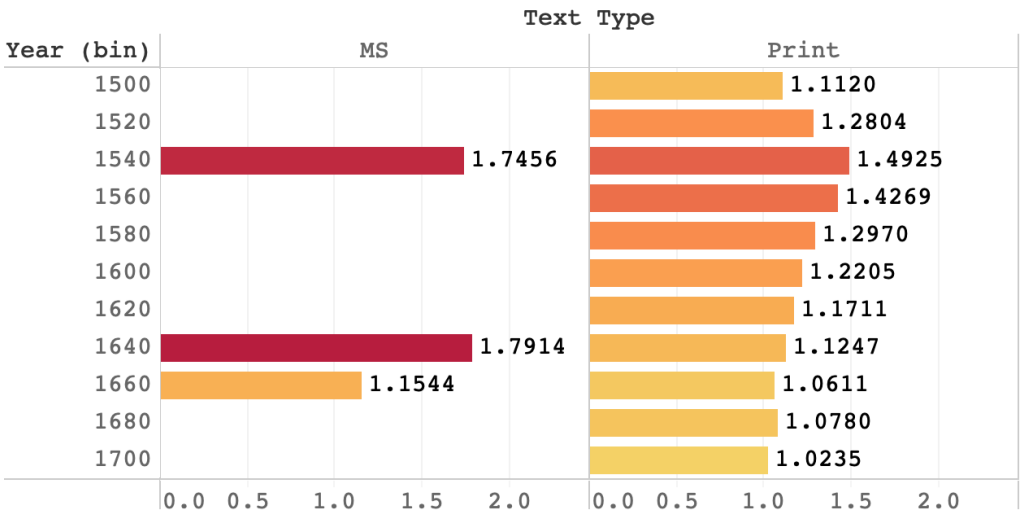

Figure 3. Aggregate mean frequencies of spelling variants in the printed texts and manuscripts.

At the extreme end of the scale, a text could feature as many as 27 unique spellings for the same lexical item. This was found in Hunter 64 and the word *betony*, today also known as the common hedgenettle, which was spelled *betaine*, *betanie*, *betanni*, *betannye*, *betany*, *betanye*, *betayne*, *betein*, *betenie*, *betenny*, *beteyne*, *betoine*, *betoni*, *betonie*, *betonye*, *bettayne*, *bettin*, *betton*, *bettoney*, *bettoni*, *bettonie*, *bettony*, *bitannie*, *bitonie*, *bitten*, *bytanie*, *bytannie* and *petanye.* The same manuscript gives 12 variant spellings of *linseed*, 10 of *celandine*, 8 of *always*, etc. Of all the texts in the two corpora, while there are some texts that show lower overall percentages of standardised spellings, Hunter 64 is by far the most varied when it comes to multiple spellings of the same word.

5.2. Lexical diversity and similarity

When VoC-D is calculated for each text in EMEMT and MCEMESP, we can first confirm that neither the year of publication nor length of text appear to correlate significantly with lexical richness. Figure 4 shows a contour plot showing the correlation between time (horizontal axis) and VoC-D (vertical axis), with the coloured areas highlighting the standardisation percentage (blue indicating low and red high percentages).

Print ISSN: 1578–7044; Online ISSN: 1989–6131

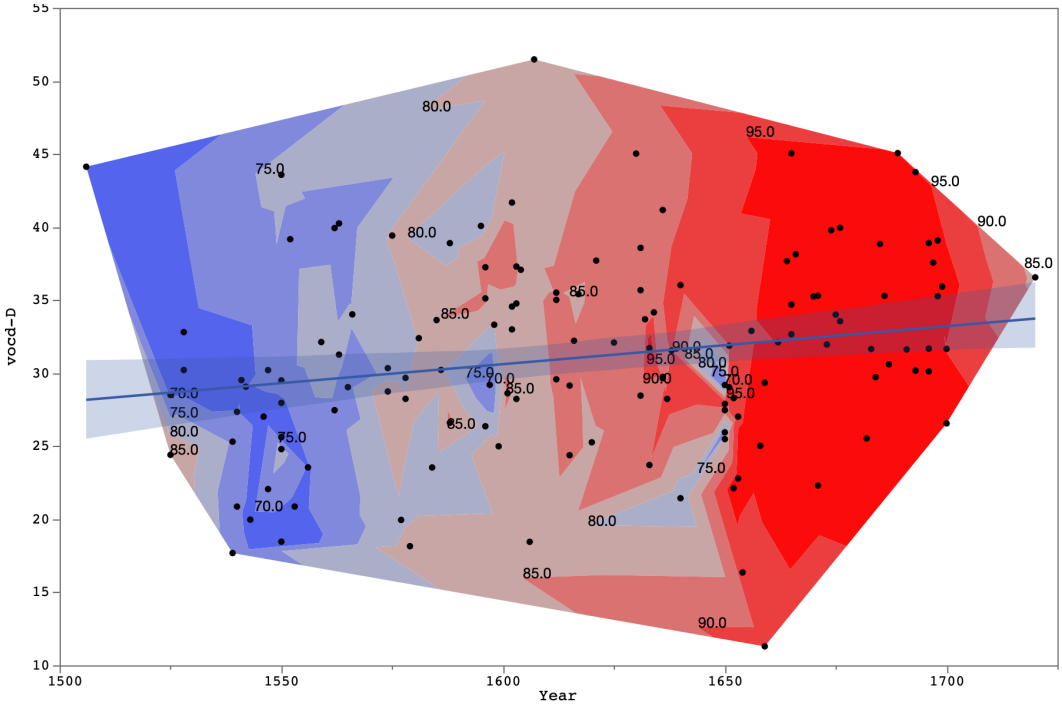

Figure 4. Correlation between time and vocabulary density.

How should we interpret this result? As the process of standardisation progressed and the number of variants spellings decreased, medical writing was at the same time getting more and more diverse in terms of lexis, at least partly in reflection of the growing professionalism and the development of a specialised medical register. It is well attested that medical writing has always been a field with a great affinity for learned vocabulary, and numerous studies have investigated the introduction of neologisms, inkhorn terms and Latinate loans during the Early Modern period (see e.g. McConchie, 1997, Norri, 2016 and Tyrkkö, 2011). However, what is noteworthy here is the co-occurrence of the two phenomena, namely, that in this specific genre of writing, it appears that a curious sort of balance was more or less maintained for two centuries when it comes to lexical diversity on the medical page.

The contour plot in Figure 5 shows the diachronic development of the relationship between the proportion of non-standard spellings, Levenshtein distance (left) and VoC-D (right).[8] As time passes, and the number of non-standard spellings decreases, the remaining non-standard words become less and less non-standard, and the vocabulary density of the texts increases.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

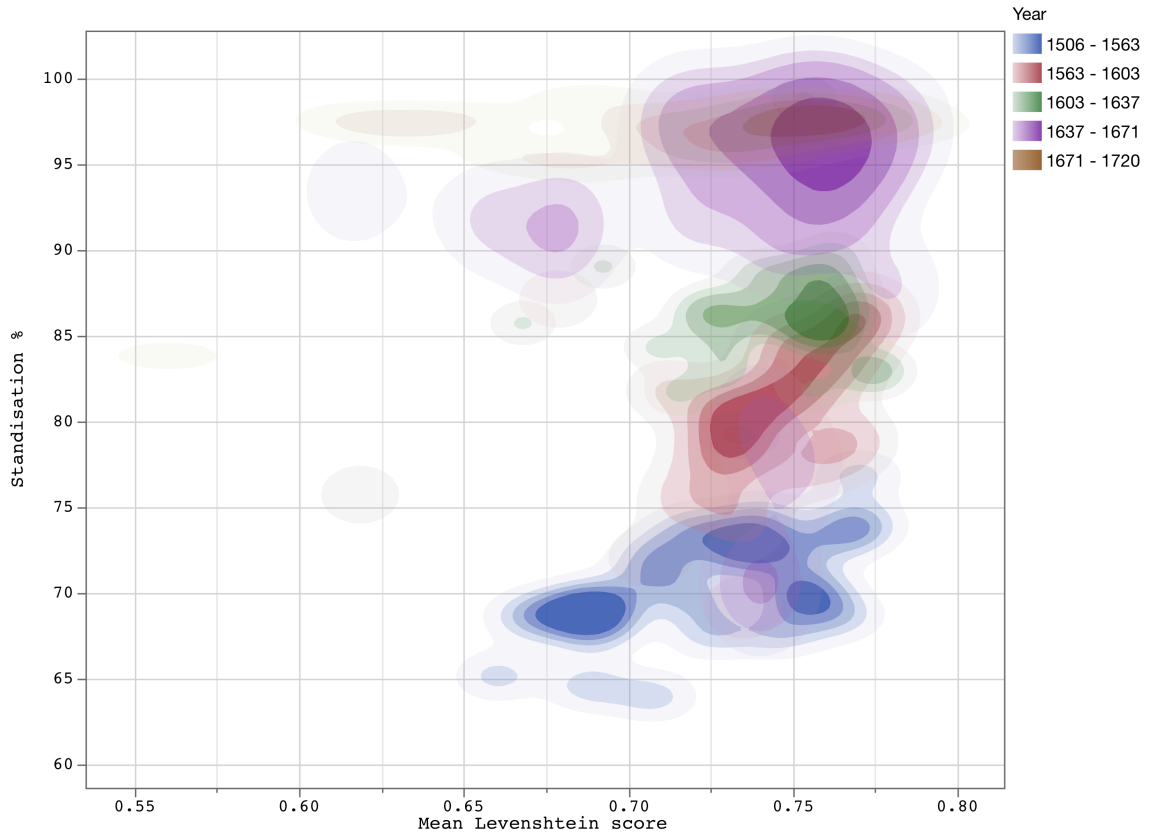

Figure 5. Plot of correlation between the proportion of standardisation, edit distance and time period.

5.3. Word-final trigrams

The general trends observed using the output of VARD standardisation and the two statistical metrics of Levenshtein distance and VoC-D are naturally only indexical of the complex orthographical developments that took place in manuscript and print texts over the two centuries covered. It goes beyond the scope of this paper to explore changes in the full range of individual features, but since it is always useful to look at actual linguistic features, the final part of the study will exemplify the distributional trends of individual spelling features by turning to word-final trigrams. Defined as the three last characters at the end of individual lexical items, these trigrams are not morphological units as such, but taken as a whole they reveal the trends of diachronic change and allow us to pinpoint the moments when specific features were replaced by their standard variants.

All in all, 4,997 unique types of word-final trigrams were discovered in the corpora in words longer than three characters. This may sound like much, but given the number of possible permutations, the theoretical maximum is in fact much greater still. In terms of counts of occurrence, the majority of these items occurred only very rarely. In order to exemplify the distributional trends, a select few higher frequency trigrams will be presented below as scatter

Print ISSN: 1578–7044; Online ISSN: 1989–6131

plots with a polynomial trendlines; the manuscript texts are indicated with square symbols in the scatter plots. It is worth noting that word-final trigrams are naturally sensitive to code-switching, which in the case of Early Modern English medical writing is especially relevant when it comes to words, phrases and longer passages written in Latin and Greek (see e.g. Honkapohja, 2018 and Taavitsainen, 2000). Consequently, the examples presented in this section were selected from among common English word-endings. Non-parametric density curves have been added to help visualise periods during which we can see particular uniformity in the frequency data, with red regions on the plot showing these areas, while yellow, green and blue regions indicate increasingly diverse usage. Typically, this would mean that a given trigram is found in a great number of different lexical items.

Figure 6 illustrates the decline in the frequency of the word-final <rte>, which is naturally only one of many forms that reflect the drop of the word-final <e>. In medical writing, <rte> occurs particularly often in types such as "parte", "hurte" and "herte". As the plot shows, the e-final form persists relatively consistently until the 1600, at which point it appears to disappear very rapidly or, in other words, reaches the second plateau of a reverse S-curve pattern.[9] Notably, the only texts that show <rte> after 1600 are the manuscript texts (indicated with square markers).

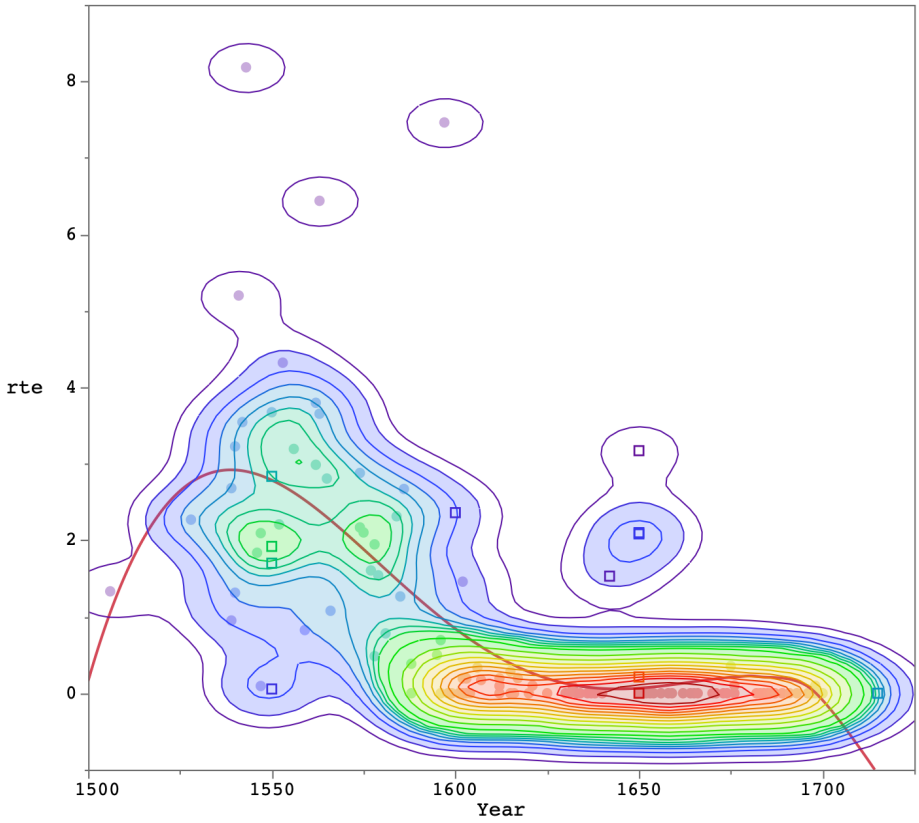

Figure 6. Plot of word-final <rte> against time.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

The opposite pattern can be observed with <ity> (Figure 7), which was of course a highly productive suffix during this time (see Säily, 2014).[10] The frequency of <ity> follows a shallow S-curve, beginning to appear during the late sixteenth century and undergoing a period of rapid adoption during the seventeenth. Notably, nearly all the manuscripts show low frequencies of <ity> during the seventeenth century.

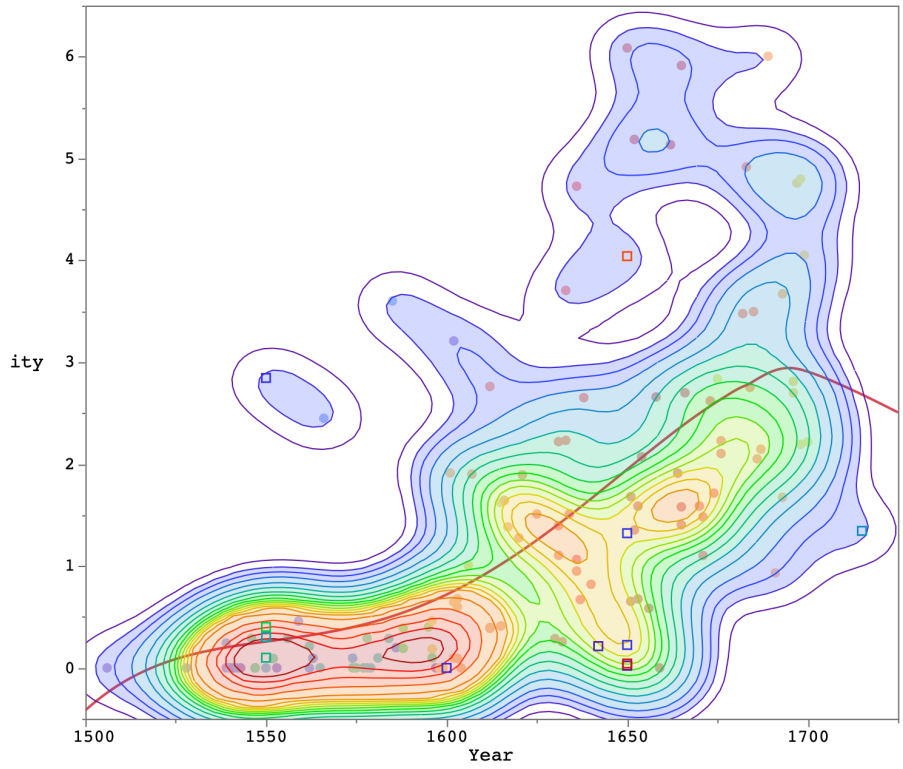

Figure 7. Plot of word-final <ity> against time.

However, in order to examine use of <ity> from the spelling standardisation perspective, we may wish to study its frequency profile in comparison to the corresponding pre-standard form, namely <itie>. Since the trigram data give access to <tie>, we use it as a proxy for comparing <tie> with <ity>. As Figure 8 shows, the frequency plots of <tie> and <ity> cross at approximately 1645.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

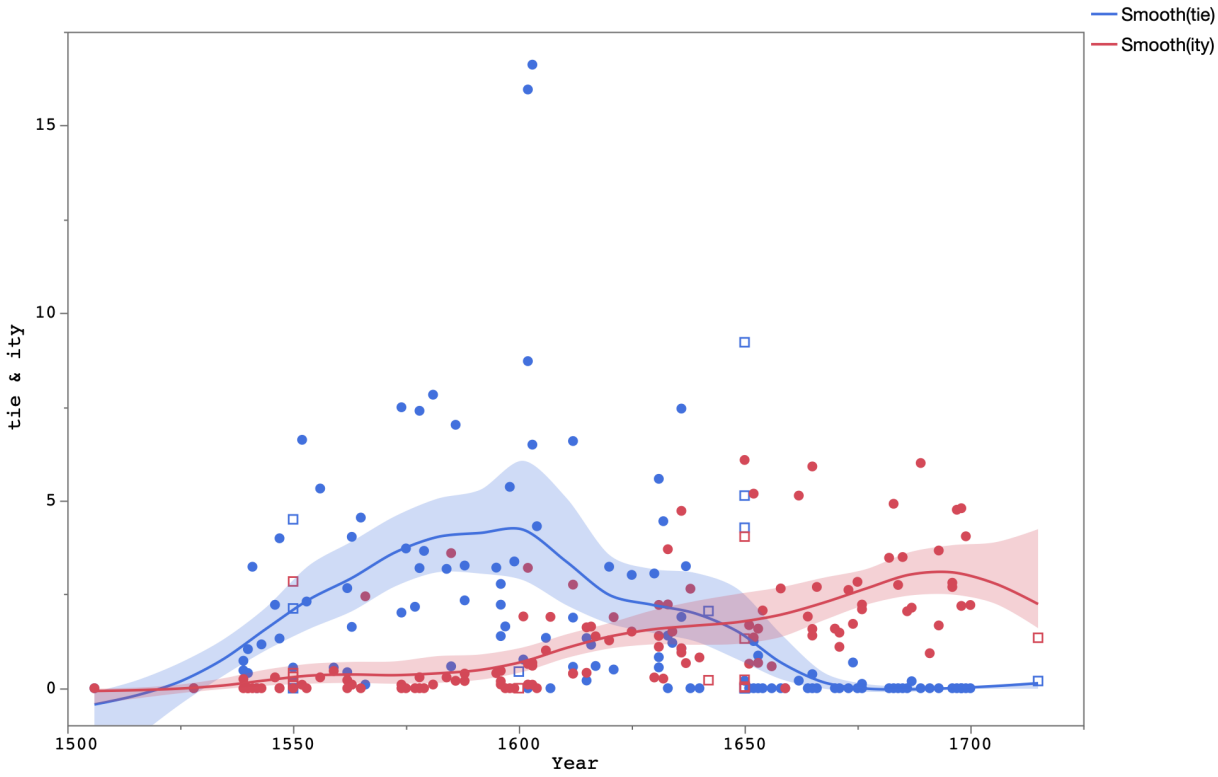

Figure 8. Plot of word-final <ity> and <tie> against time.

Moving on, Figure 9 shows the frequency pattern of <irm>, which only appears after 1650 but is then found in more than 20 texts in the corpora. Here again the pattern is explained by the disappearance of the word-final <e>, with word forms such as "confirme" and "affirme" disappearing and being replaced by the standard variants "confirm" and "affirm". The corpus evidence suggests that in this case the switch happens suddenly, not gradually as we might expect considering that the texts were written, typeset and printed by numerous different individuals. This observation is partly explained by the fact that the range of words ending in <irm> is relatively low.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

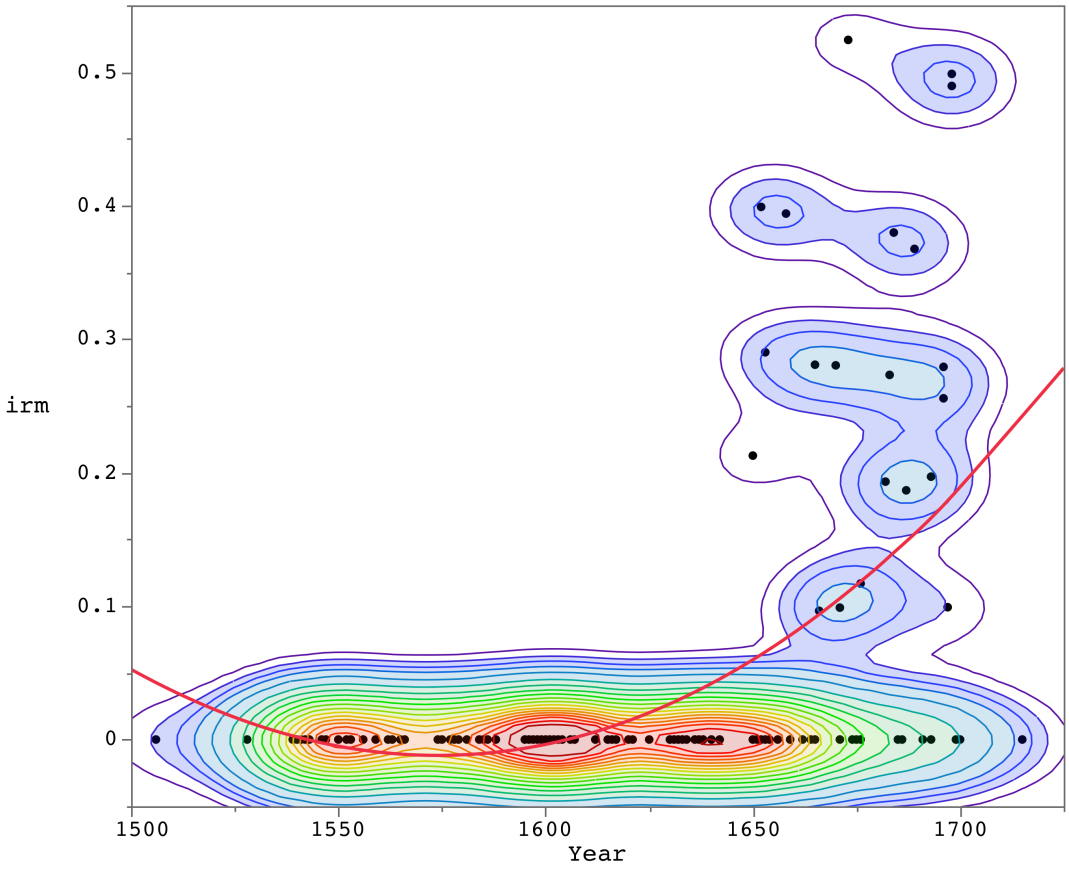

Figure 9. Plot of word-final <irm> against time.

The comparison of these and other frequency plots of corresponding non-standard and standard trigrams points to the early seventeenth century as the turning point. In Figure 10, we see that the plots of the word-final forms <uer> and <ver> cross around 1625, or approximately 20 years later than the <tie> to <ity> switch seen earlier. This date agrees extremely well with previous scholarship and the dating of the height of the English Spelling Reform; see e.g. Scholfield (2016: 149), who gives the date for the establishment of the lower case <u> and <v> as distinct letters as 1630. What is notable here is that there are hardly any occurrences of the non-standard <uer> form after that point and all of those are found in the manuscripts. We can examine the feature further by casting each text as a point in a scatter plot, which shows that very few of the texts exhibit internal variation between the <uer> and <ver> endings. Of the 14 texts that do, seven are manuscripts, which is a significantly high proportion given the overall numbers of texts (7/12 vs. 7/121).

Print ISSN: 1578–7044; Online ISSN: 1989–6131

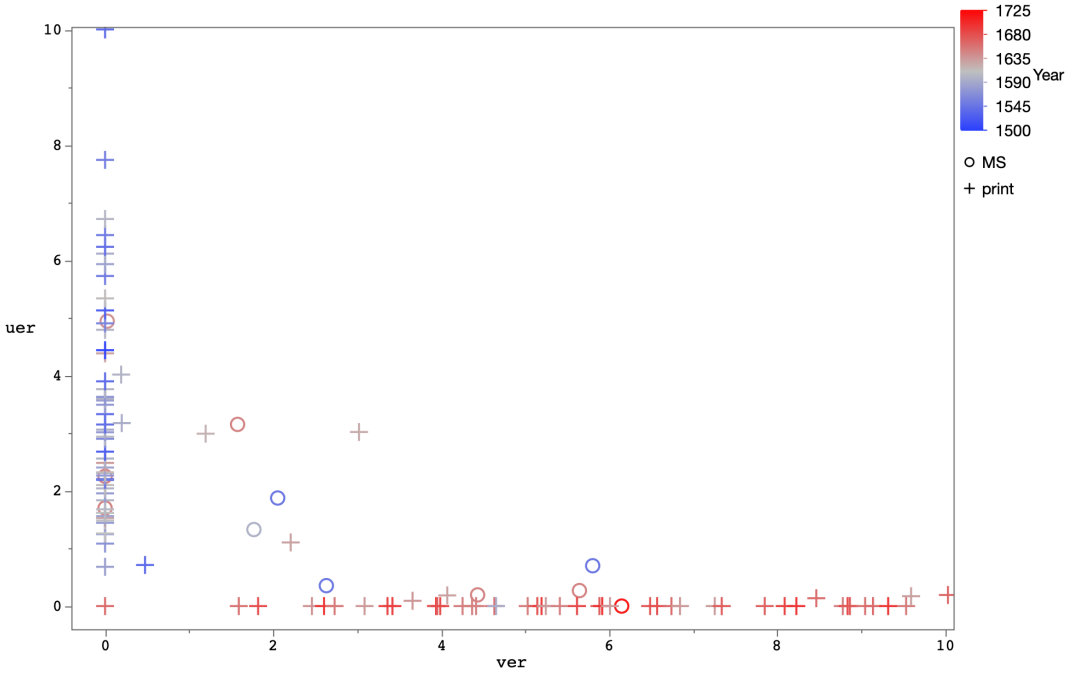

Figure 10. Plot of word-final <uer> and <ver> against time.

Our final example is <yre>, which presents two successive waves of standardisation (Figure 11). Up until 1550, "ayre" is the only option, from 1550 to around 1660 "aire" is the dominant form with "ayre" persisting to some extent, and around 1660 "aire" is replaced by "air" as the final <e> is dropped.[11] Again the use of visualisations helps us understand the progression of the process in an approachable way, and also allows us to see the existence of outliers. Note that the clines seen in the trendlines after 1700 are merely artefacts of the fitting technique, they do not indicate that <ire> would have experienced a sudden increase at the end of the period.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

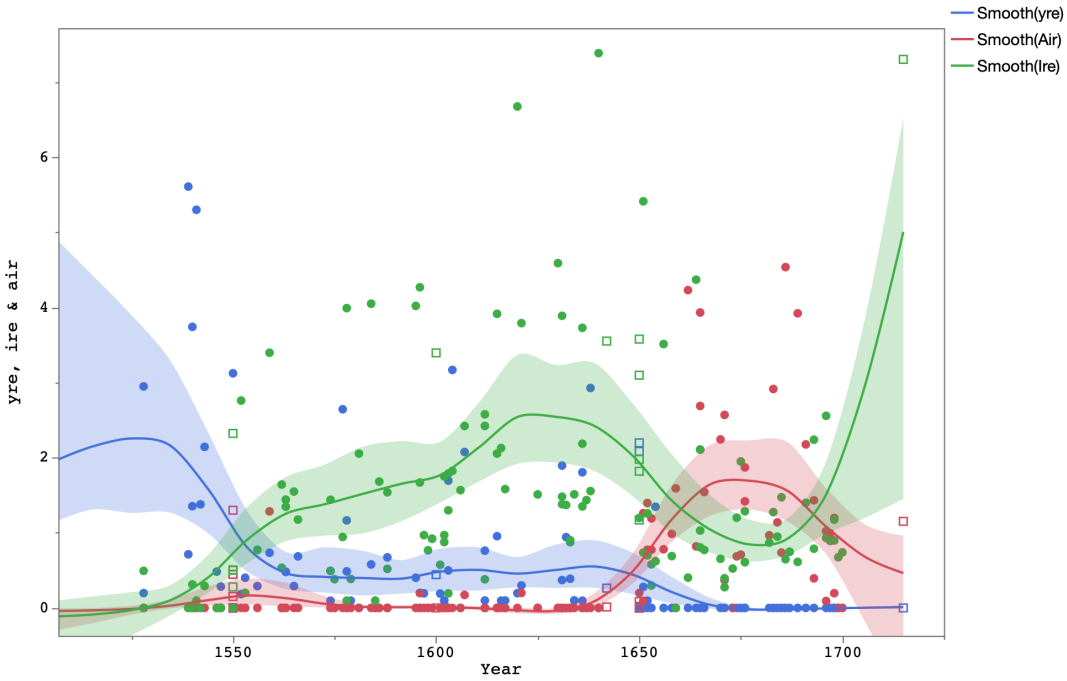

Figure 11. Plot of word-final <yre> against time.

7. CONCLUSIONS

This paper has introduced a selection of descriptive and largely quantitative methods for approaching spelling standardisation during the Early Modern period. The findings were presented using a variety of data visualisation methods which make the trajectories of the processes considerable easier to understand than tabular representations, which more often than not bin together data using arbitrary chunks of time and large numbers of individual texts without much regard for variation within each section of the corpora.

The results of the analysis largely confirmed previous knowledge of how spelling standardisation proceeded, which can be considered evidence of the fact that the retrieval of relevant items and the quantitative analysis were correctly executed. However, there were also observations that offer new information about the spelling in Early Modern medicine. The difference between the printed and hand-written medical corpora showed that while the standardisation process proceeded along a remarkably steady path in the printed texts, showing no apparent outliers among a group of nearly 100 printing houses, the manuscripts offered a much more varied picture of individual variation. While the greater variety in manuscripts is not in itself particularly noteworthy, the fact that these texts belong to a developing professional register suggests that spelling was not a topic of particular concern for the medical professionals or their scribes. If we contrast this with present-day academic writing, one of the fundamental requirements in most contexts is that the language is free of errors and the spelling

Print ISSN: 1578–7044; Online ISSN: 1989–6131

is correct. Given the timelines seen in Figure 1, while printed medical texts were nearly fully standardised by the end of the Early Modern period, contemporary manuscripts were at a level that was nearly a century behind the printed texts. The manuscripts also showed that the range of variation in spelling standardisation was remarkably wide, ranging from texts that were entirely on par with printed books to others that were strikingly diverse, even when it came to the spelling of individual words.

ACKNOWLEDGEMENTS

The author is a long-time member of the Scientific Thought-styles project at the University of Helsinki and a member of the advisory board of *The Málaga Corpus of Early English Prose* at the University of Málaga. The author is grateful for the insightful comments received during the refereeing process.

NOTES

1 EMEMT is the second part of the three-part *Corpus of Early English Medical Writing*, compiled at the University of Helsinki by the Scientific Thought-styles project, led by Prof. Irma Taavitsainen.

2 The term 'category' was adopted by the compilers in order to avoid theoretically loaded terms such as genre, register or text type. The categories in EMEMT are somewhat heuristically determined, some reflecting topic while others are more closely related to medical discipline or text type (such as periodicals); see Taavitsainen (2010).

3 English printers imported fonts from the continent, where thorn was not used (see e.g. Hill, 2016). A rare example of a font cut in England featuring Anglo-Saxon alphabet was the one created by John Day for Archbishop Matthew Parker in 1566; see Hotchkiss and Robinson (2008: 10).

4 MCEMESP is a part of *The Málaga Corpus of Early English Prose*. The project is led by Prof. Javier Calle-Martín.

5 For latest versions of VARD, see http://ucrel.lancs.ac.uk/vard/versions/. A separate tool for analyzing VARD output called DICER also exists, but it was not used in the present study; see Baron et al. (2011).

6 The quality of VARD's performance is extensively discussed in Baron (2011) and Baron et al. (2011). The scope of the present study prevents a more detailed analysis and consequently the results of VARD were used without further examination.

7 The summary statistics for the two linear trend lines are r=0.95, p=***, R^2=0.91 for the print texts and r=0.74, p=*, R^2=0.54 for the manuscripts.

8 The contour plot is based on Delaunay triangulation, where triangulation and linear interpolation are used to generate the contours.

9 Note that the rising cline at the beginning of the curve reflects the result of one very early text (John Lydgate's *Gouernall Of Helthe*, 1506).

Print ISSN: 1578–7044; Online ISSN: 1989–6131

10 As discussed in Hiltunen and Tyrkkö (2011), although the increase of nominalisations in scientific writing has been associated with the eighteenth century, the process started earlier.

11 Note that not all instances of word-final <yre> are of parts of the longer 4-gram <ayre>; the corpus attests items such as *myre*, *heyre*, etc.

REFERENCES

Baron, A. (2011). *Dealing with Spelling Variation in Early Modern English Texts*. Unpublished Doctoral Dissertation, Lancaster University, United Kingdom.

Baron, A., Rayson, P. & Archer, D. (2011). Innovators of Early Modern English spelling change: Using DICER to investigate spelling variation trends. Paper presented at the *Helsinki Corpus Festival*. University of Helsinki, Finland, September–October, 28–2.

Bland, M. (2010). *A Guide to Early Printed Books and Manuscripts*. London: Wiley-Blackwell.

Boffey, J. (2012). *Manuscripts and Print in London c. 1475–1530*. London: The British Library.

Brown, R. D. (2013). Selecting and weighting n-grams to identify 1100 languages. In I. Habernal & V. Matoušek (Eds.), *Text, Speech, and Dialogue: 16th International Conference, TSD 2013. Pilsen, Czech Republic, September 2013. Proceedings* (pp. 475–483). Berlin: Springer Verlag.

Calle-Martín, J., Moreno-Olalla, D., Esteban-Segura, L., Marqués-Aguado, T., Romero-Barranco, J., Thaisen, J. & Rutkowska, H. (2016–). *The Málaga Corpus of Early Modern English Scientific Prose*. Málaga: University of Málaga. Available from https://modernmss.uma.es

Doval, S. (1996). The English spelling reform in the light of the works of Richard Mulcaster and John Hart. *Sederi, 7,* 115–126.

Eisenstein, E. L. (1979). *The Printing Press as an Agent of Change. Communications and Cultural Transformations in Early-Modern Europe. Volumes I and II*. Cambridge: Cambridge University Press.

Grafton, A. (2012). *The Culture of Correction in Renaissance Europe*. London: The British Library.

Hämäläinen, M., Säily, T., Rueter, J., Tiedemann, J. & Mäkelä, E. (2018). Normalizing early English letters to Present-day English spelling. In B. Alex, S. Degaetano-Ortlieb, A. Feldman, A. Kazantseva, N. Reiter & S. Szpakowicz (Eds.), *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (pp. 87–96). Stroudsburg, PA: ACL.

Heikkilä, T. & Roos, T. (2018). Quantitative methods for the analysis of medieval calendars. *Digital Scholarship in the Humanities, 33*(4), 766–787.

Hill, W. (2016). Typography and the printed English text. In V. Cook & D. Ryan (Eds.), *The Routledge Handbook of the English Writing System* (pp. 431–452). Abingdon: Routledge.

Hiltunen, T. & Tyrkkö, J. (2009). Frequency of nominalization in Early Modern English medical writing. In A. Jucker, M. Hundt & D. Schreier (Eds.), *Corpora: Pragmatics and Discourse. Papers from the 29th International Conference on English Language Research on Computerized Corpora* (pp. 293–316). Amsterdam: Rodopi.

Honkapohja, A. (2018). "Latin in Recipes?" A corpus approach to scribal abbreviations in 15th-century medical manuscripts. In P. Pahta, J. Skaffari & L. Wright (Eds.), *Multilingual Practices in Language History: English and Beyond* (pp. 243–271). Berlin: De Gruyter.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

Honkapohja, A. & Liira, A. (2020). Abbreviations and standardisation in the *Polychronicon*: Latin to English, and manuscript to print. In L. Wright (Ed.), *The Multilingual Origins of Standard English* (pp. 269-316). Berlin: Mouton de Gruyter.

Hotchkiss, V. & Robinson, F. C. (2008). *English in Print from Caxton to Shakespeare to Milton*. Urbana, IL & Chicago, IL: University of Illinois Press.

Howard-Hill, T. H. (2006). Early modern printers and the standardization of English spelling. *The Modern Language Review, 101*(1), 16–29.

Kestemont, M. (2015). A computational analysis of the scribal profiles in two of the oldest manuscripts of Hadewijch's letters. *Scriptorium, 69*(2), 159–177.

Lehto A., Baron, A. & Rayson, P. (2010). Improving the precision of corpus methods: The standardized version of Early Modern English Medical Texts. In I. Taavitsainen & P. Pahta (Eds.), *Early Modern English Medical Texts: Corpus Description and Studies* (pp. 279–290). Amsterdam: John Benjamins.

Marotti, A. F. (1995). *Manuscript, Print, and the English Renaissance Lyric*. Ithaca, NY: Cornell University Press.

Maruca, L. (2003). Bodies of Type: The work of textual production in English printers' manuals. *Eighteenth-century Studies, 36*(3), 321–343.

McConchie, R. W. (1997). *Lexicography and Physicke: The Record of Sixteenth-century English Medical Terminology*. Oxford: Clarendon Press.

McKee, G. T., Richards, B. & Malvern, D. (2000). Measuring vocabulary diversity using sedicated software. *Literary and Linguistic Computing, 15*(3), 323–337.

McKerrow, R. B. (1967). *An Introduction to Bibliography for Literary Students*. Oxford: Clarendon Press.

Moessner, L. (2017). Standardization. In A. Bergs & L. J. Brinton (Eds.), *The History of English. Volume 4: Early Modern English* (pp. 167–187). Berlin: Mouton de Gruyter.

Moxon, J. (1683). *Mechanick Exercises*. London: Printed for Joseph Moxon on the west-side of Fleet-ditch.

Neumann, J. H. (1944). Jonathan Swift and English spelling. *Studies in Philology, 41*(1), 79–85.

Nevalainen, T. (2012). English: Variable focusing in English spelling between 1400 and 1600. In S. Baddeley & A. Voeste (Eds.), *Orthographies in Early Modern Europe* (pp. 127–165). Berlin: Mouton de Gruyter.

Nevalainen, T. & Tieken-Boon van Ostade, I. (2006). Standardisation. In R. Hogg & D. Denison (Eds.), *A History of the English Language* (pp. 271–311). Cambridge: Cambridge University Press.

Norri, J. (2016). *Dictionary of Medical Vocabulary in English, 1375–1550: Body Parts, Sicknesses, Instruments, and Medicinal Preparations*. London: Routledge.

Osselton, N. (1984). Informal spelling styles in Early Modern English: 1500–1800. In N. F. Blake & C. Jones (Eds.), *English Historical Linguistics*. *Studies in Development* (pp. 123–137). Sheffield: Department of English Language, University of Sheffield.

Pahta, P. (2011). Code-switching in Early Modern English medical writing. In I. Taavitsainen & P. Pahta (Eds.), *Medical Writing in Early Modern English* (pp. 115–134). Cambridge: Cambridge University Press.

Print ISSN: 1578–7044; Online ISSN: 1989–6131

Pahta, P., Hiltunen, T., Marttila, V., Ratia, M., Suhr, C. & Tyrkkö, J. (2011). Communicating Galen's *Methodus medendi* in Middle and Early Modern English. In P. Pahta &. A. H. Jucker (Eds.), *Communicating Early English Manuscripts*. Studies in English Language (pp. 178–196). Cambridge: Cambridge University Press.

Pettersson, E., Megyesi, B. & Nivre, J. (2013). Normalisation of historical text using context-sensitive weighted Levenshtein distance and compound splitting. In *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013)* (Linköping Electronic Conference Proceedings no 85) (pp. 163–179). Linköping: Linköping University Electronic Press.

Romero-Barranco, J. (2017). *Early Modern English Scientific Text Types: Edition and Assessment of Linguistic Complexity of the Texts in MS Hunter 135 (ff. 34r–121v).* Unpublished Doctoral Dissertation, University of Málaga, Spain.

Rutkowska, H. (2013). Towards regularisation: Morphological spelling in several editions of the *Kalender of Shepherdes*. *Studia Anglica Posnaniensia, 48*(1), 7–28.

Rutkowska, H. (2016). Orthographic regularization in Early Modern English printed books: Grapheme distribution and vowel length indication. In C. Russi (Ed.), *Current Trends in Historical Sociolinguistics* (pp. 165–193). Berlin: De Gruyter Open.

Rutkowska, H. (2017). Orthography. In L. J. Brinton & A. Bergs (Eds.), *Historical Outlines from Sound to Text* (pp. 200–217). Berlin: De Gruyter.

Säily, T. (2014). *Sociolinguistic Variation in English Derivational Productivity: Studies and Methods in Diachronic Corpus Linguistics*. Helsinki: Société Néophilologique de Helsinki.

Sairio, A., Kaislaniemi, S., Merikallio, A. & Nevalainen, T. (2018). Charting orthographical reliability in a corpus of English historical letters. *ICAME Journal, 42,* 79–96.

Scholfield, P. (2016). Modernization and standardisation since the seventeenth century. In V. Cook & D. Ryan (Eds.), *The Routledge Handbook of the English Writing System* (pp. 143–165). London: Routledge.

Scragg, D. G. (1974). *A History of English Spelling*. Manchester: Manchester University Press.

Sönmez, M. J.-M. (1993). *English Spelling in the Seventeenth Century: A Study of the Nature of Standardisation as Seen through the MS and Printed Versions of the Duke of Newcastle's 'A New Method…'*. Unpublished Doctoral Dissertation, University of Durham, United Kingdom.

Stenroos, M. & Smith, J. (2016). Changing functions: English spelling before 1600. In V. Cook & D. Ryan (Eds.), *The Routledge Handbook of the English Writing System* (pp. 125–142). London: Routledge.

Taavitsainen, I. (2000). Scientific language and spelling standardization 1375–1550. In L. Wright (Ed.), *The Development of Standard English, 1300–1800: Theories, Descriptions, Conflicts* (pp. 131–154). Cambridge: Cambridge University Press.

Taavitsainen, I., Murray Jones, P., Pahta, P., Hiltunen, T., Marttila, V., Ratia, M., Suhr, C. & Tyrkkö, J. (2011). Medical texts in 1500–1700 and the corpus of Early Modern English Medical Texts. In I. Taavitsainen & P. Pahta (Eds.), *Medical Writing in Early Modern English* (pp. 9–29). Cambridge: Cambridge University Press.

Taavitsainen, I., Pahta, P., Hiltunen, T., Mäkinen, M., Marttila, V., Ratia, M., Suhr, C. & Tyrkkö, J., (Comps.). (2010). *Early Modern English Medical Texts* (CD-ROM). Amsterdam: John Benjamins.

Tieken-Boon van Ostade, I. (1998). Standardisation of English spelling: The eighteenth-century printers' contribution. In J. Fisiak & M. Krygier (Eds.), *Advances in English Historical Linguistics (1996)* (pp. 457–470). Berlin: Mouton de Gruyter.

Tyrkkö, J. (2011). 'Halles Lanfranke' and its most excellent and learned expositive table. In O. Timofeeva & T. Säily (Eds.), *Words in Dictionaries and History: Essays in Honour of R. W. McConchie* (pp. 17–39). Amsterdam: John Benjamins.

Tyrkkö, J. (2013). Printing houses as communities of practice: Orthography in early modern medical books. In J. Kopaczyk & A. H. Jucker (Eds.), *Communities of Practice in the History of English* (pp. 151–176). Amsterdam: John Benjamins.

Tyrkkö, J., Hickey, R. & Marttila, V. (2010). Exploring Early Modern English medical texts: Manual to EMEMT presenter. In I. Taavitsainen & P. Pahta (Eds.), *Early Modern English Medical Texts: Corpus Description and Studies* (pp. 221–279). Amsterdam & Philadelphia, PA: John Benjamins.

Print ISSN: 1578–7044; Online ISSN: 1989–6131