Towards new knowledge: The corpus of *Late Modern English Medical Texts*

Turo Hiltunen and Irma Taavitsainen

Abstract

Late Modern English Medical Texts (LMEMT) is a new corpus representing printed medical writing in the eighteenth century. This chapter describes the structure and the main compilation principles of the corpus. Representativeness is a complex notion in corpus linguistics in general, and the issue is particularly challenging in the context of eighteenth-century medicine, where the volume of published texts increased considerable and the scope of the discipline widened. To provide a realistic picture of the variety of medical texts in this century, the field is divided into text categories, which reflect contemporary divisions and incorporate texts written for different purposes and addressed to different audiences. The corpus is designed for studies in areas such as linguistics, pragmatics, medical history, and digital humanities, which are showcased in the contributions to this book.

> *[T]o act laudably and gain success is the triumph of wisdom; it is the way to be acquainted with nature.*

> (Anderson, *A preliminary introduction into the act of sea-bathing*, 1795: 27)

1. Introduction

Historical corpus studies rely on real language use of the past as recorded in digital corpora. Studies on historical corpora can be either synchronic (focusing on a specific historical period) or diachronic (tracing patterns of variation and change over a given time period). What these perspectives have in common is the reliance on empirical assessments of authentic language use, as represented in systematically compiled text collections. The well-known division in the general and specific corpora (e.g. McEnery et al. 2006) also applies to diachronic corpora: general corpora are designed for the study of the language variety as a whole, whereas specialized corpora allow

the study of individual sub-languages, conceptualized in terms of genres, registers, modes, topics, categories, and the like. The corpus of *Late Modern English Medical Texts* (LMEMT) is a specialized diachronic corpus, representing medical writing between 1700 and 1800, providing comprehensive data for analysis. The corpus has been designed with different user groups in mind: while primarily meant for linguistic study, it can be of assistance to scholars of other disciplines, since text passages can easily be located with corpus searches.

Late Modern English Medical Texts (LMEMT) is the third corpus of a series that began with *Middle English Medical Texts* (MEMT, 2005) and continued with *Early Modern Medical Texts* (EMEMT, 2010) 1500–1700. Together these corpora provide material for long-term chronological studies that cover more than four centuries. Our aim in this volume is to demonstrate the potential of the new resource and give inspiration to new explorations on the corpus materials. A great abundance of medical texts is extant from the eighteenth century, and thus the compilation task was even more challenging than before. Our solution has been to collaborate with medical historians of the Faculty of Early Science and Medicine at the University of Cambridge to ensure the best possible result.

The motivation for undertaking the task of compiling the three corpora of historical medical writing in English was the desire to discover patterns of variability and draw the main lines of development in more detail than has been done before in this important specialized field of writing. Medicine was the spearhead field within scientific writing for centuries and even today it shows conventionalized features, such as the use of complex noun phrases and reduced verbs, perhaps to a greater extent than any other discipline, as it is used by the worldwide discourse community for communicating new knowledge.

2. The corpus of *Late Modern English Medical Texts*

LMEMT is the third register-specific corpus, containing over two million words from a wide range of eighteenth-century medical texts. Significant changes were taking place in this period both in the underlying philosophy of science as well as in the ways of disseminating and communicating medical information. These changes are reflected in the corpus texts and provide plenty of uncharted materials. Studies on

their linguistic features can yield more detailed knowledge of the developments, and the long diachronic line opens up new possibilities for exploration.

The three corpora are designed to be maximally compatible with one another in structural terms, although due to changes in time full parallelism is not attainable. The structure of the LMEMT corpus has been carefully planned to provide continuation to EMEMT and also MEMT. The main categories of late medieval medical writing – SPECIALIZED TEXTS, SURGICAL TEXTS and REMEDIES – continue all through the three corpora, with an increasing number of texts written on these topics. Alongside these fields, new categories of medical texts have emerged over the following centuries, including institutional writing in the category of PUBLIC HEALTH. Figure 1 provides a schematic representation of the categories and their relationship.
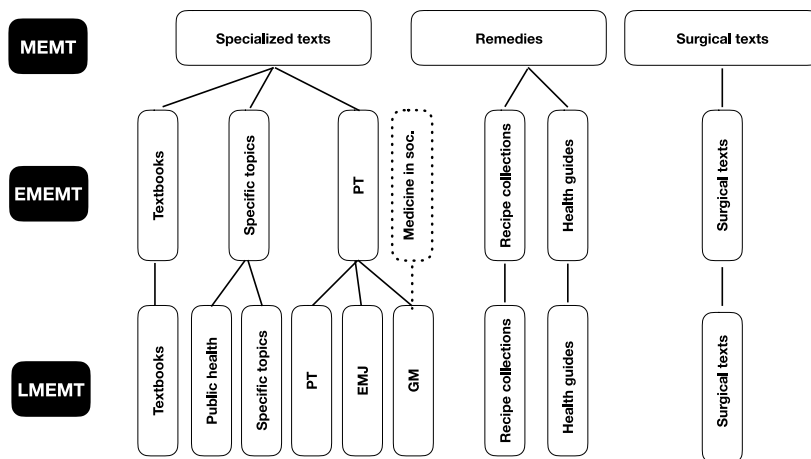

Figure 1. Text categories across three corpora

Overall, the distribution of the texts over the 100-year time period is relatively even, which was our aim, but the individual categories are not fully balanced chronologically. The selected texts are meant to reflect the contemporary textual reality as closely as possible, and we have accordingly consulted statistics on printed books where those have been available. However, text selection is also influenced by what titles were available in an electronic format at the time of compilation. Decisions on what texts to include and what to leave out were made based on an extensive survey of all medical titles on ECCO, secondary literature on eighteenth-century medical authors and their works,[1] and finally consultation with Dr. Peter Jones. The distribution of the texts into the categories is shown below in figures 2–8 below, and the chronological coverage of each text category and the principles on which the texts were chosen are discussed in more detail the Section of category descriptions.

With slightly over two million words of a wide range of medical texts, LMEMT is only moderately large compared to modern mega-corpora (see e.g. Hiltunen et al. 2017 for a discussion), but its moderate size is offset by our careful and principled text selection, detailed description, and contextualization of the texts. At the same time, compared to other specialized historical corpora, LMEMT is not so small.[2] All in all, it clearly represents what Mair (2006: 355) has termed "a small-and-tidy approach" in its emphasis on careful text selection and laborious manual annotation. Some conventions adopted for the earlier corpora continue in LMEMT, but there are also innovations such as the XML mark up (see the Manual in this volume).

The aim of our corpus compilation has been to provide as representative a sample as possible of printed medical texts between 1700 and 1800, selected according to text-external criteria. For our selection of texts, we have adopted an inclusive view of medicine that covers the whole domain of medical writing, ranging

---

[1] Notably the *Oxford Dictionary of National Biography* (ODNB).

[2] E.g. the different sections of the *Coruña Corpus of Scientific Writing* (Moskowich et al. 2012, 2016) contain roughly 400,000 words each. ARCHER (A Representative Corpus of Historical English Registers) includes medical texts as well, but the selection is very concise, as ARCHER is a multi-genre historical corpus of British and American English covering the period between 1600 and 1999.

from texts written by and targeted at the academic elite to treatises for household use, recipes, advertisements and writings intended for semiliterate audiences. The production and circulation of medical texts show some new developments, e.g. novel channels of publication in the form of specialized medical journals for learned professionals and communication about medical issues more broadly in another innovative medium, *the Gentleman's Magazine*. In general, the majority of corpus texts were written by learned doctors for their peers, but the corpus also contains medical texts aimed at more heterogeneous groups of lay people, aiming to cover what is extant and what was available. The corpus covers the full range of printed medical writing in the late modern period, with its rich diversity, and is divided into eight text categories to facilitate studies on different sub-registers of the field (see below).

3. What LMEMT represents

To enable systematic studies on the diverse trends and developments in eighteenth-century medical writing, a comprehensive and as representative a collection of medical texts as possible is clearly a necessity. To this end, the LMEMT corpus has been compiled to be representative of printed medical texts in English in the eighteenth century. While this aim is straightforward in itself, there are a number of issues that need to be considered. Indeed, representativeness is a key concept in corpus linguistics, and one that is present in all corpus compilation projects, at least implicitly. The issue has also received substantial attention discussed in the corpus linguistic literature (see e.g. Biber 1993; Sinclair 2005; McEnery et al. 2006). It has also been pointed out that true representativeness in the statistical sense of the term is hardly attainable in corpus linguistics due to the fact that in most cases it is impossible to delimit the target population and draw a truly random sample from it (see also Evert 2006 on the problem of non-randomness in corpus linguistics in general). Even so, we agree with Sinclair's (2005) view that representativeness should remain as a target notion guiding the compilation of the corpus, and we have followed this idea in our selection of texts. Here we shall discuss the main issues pertaining to the context of eighteenth-century English medical writing and explain the solutions that we have adopted in the compilation of LMEMT. This information is extremely important for the use of the corpus, as the validity of all claims and conclusions is ultimately

predicated on the question of what the corpus actually represents. Our solutions also vary slightly between the corpus categories, and more specific information about each category is provided in the descriptions.

Most corpus linguistic studies adopt an extensional view of language, where the corpus constitutes a finite set of utterances and acts as a sample of the relevant language variety in its entirety (e.g. Baroni and Evert 2009). This is a major convenience to a researcher, because, to quote Leech's formulation, "the study of a corpus can stand proxy for the study of some entire language or variety of language" (Leech 2007: 135).

In practice, however, the goal of representativeness is elusive and difficult to achieve and there are several reasons for this. Firstly, as the above formulations suggest, compiling a corpus always involves choices: some texts are selected to form part of the corpus while other texts, which could also have been chosen, are left out. The grounds on which these choices are made are naturally subject to criticism, and it is therefore important that the selection criteria are made clear. In the case of LMEMT, for example, we want to highlight the fact that for practical reasons our target population is all printed medical texts. This is an essential piece of information for users: the learned end of the scale is well covered and as fully representative as possible, but we cannot make the same claim of the "popular" end of the scale for several reasons. Handwritten materials circulated widely and are important for the dissemination of recipes, for instance; and epistolary medical advice flourished in the eighteenth century (Leong and Pennell 2007; Wild 2006; Brown 2011), and much of the ephemeral medical data like advertisements have not survived. The most popular layers of writing that contain inherited wisdom with almanac lore and popular astrology were, however, beyond our reach, as the surviving texts still remain largely uncharted in their repositories (see notes 14–18 in Chapter 3 in this volume). Our text selection has also been guided by availability. The main source of corpus texts is the online repository *Eighteenth Century Collection Online* (ECCO), which provides access to facsimile images of eighteenth-century printed texts. Some of the texts in XML format were obtained through institutional collaboration with the ECCO Text Creation Partnership (TCP) based in Michigan. To complement the selection, a

number of texts have been obtained through agreements with various repositories, and they have been keyed in.

The task of compiling a maximally representative corpus of eighteenth-century medical writing is naturally large and challenging. To facilitate the endeavour, we have divided the field of medicine into eight broad areas, which make up the structure of LMEMT and correspond to the seven text categories and the Appendix in EMEMT. Importantly, the text categories were identified based on disciplinary factors and medical history. The criteria for choosing the texts are thus strictly text-external, focusing on the fields of medicine and the topics of texts. The categories are GENERAL TREATISES AND TEXTBOOKS, SPECIALIZED TREATISES (divided into five sub-categories), REMEDYBOOKS, REGIMEN TEXTS AND HEALTH GUIDES, SURGICAL TEXTS, PUBLIC HEALTH, which is completely new, and PERIODICALS, divided into two sub-categories (see the descriptions in this volume). In addition to *The Philosophical Transactions*, specialized medical periodicals were established in this century, and *Edinburgh Medical Journal* has been added to the category of scientific periodicals. Medical topics were also discussed more widely in the written form in newspapers and in *The Gentleman's Magazine*, which, in a sense, is a counterpart of the APPENDIX to the EMEMT corpus labelled "Medicine in Society".

We believe that our chosen approach is superior to the alternatives, because by relying on medical history and external criteria it is possible to obtain a more realistic picture of medical writing as it was perceived at the time the texts were written. This approach also facilitates the diachronic study of medicine across a longer time period than that covered by LMEMT: even if the field of medical writing becomes increasingly complex and diverse over time, the topics discussed in the texts, such as individual diseases, health advice, or childbirth, are the same in many cases. Topics therefore provide a solid basis for longer diachronic assessments of this special language. We also want to emphasize here that our text categories do not directly correspond to genres, registers, text types, or other similar variables that are commonly used in corpus linguistics. At the same time, establishing the parallels between these variables and our text categories is an exciting research task, which is possible to tackle with the help of LMEMT. In other words, we do not want research to be constrained by the text classification that we have offered, but welcome users

adjusting it to pursue different research questions, and they are encouraged to complement the data according to their research interests and expand beyond these categories to include other medical texts available in repositories like Eighteenth-century Collections Online (ECCO). Ideas for research topics and other information is provided in detail in the various category descriptions.

The underlying textual reality behind LMEMT is also reflected in the fact that the vast majority of texts are written by men (but see the MIDWIFERY description in this volume). In contrast to manuscript circulation in recipe writing, printed medical texts by female authors are fairly few, and it is even likely that the eighteenth century has proportionately fewer female authors than the seventeenth century, when household books by noblewomen flourished and female-authored midwives' manuals were prominent. The apparent over-representation of male authors in LMEMT thus reflects the actual historical situation. Text samples in each category of LMEMT aim to represent the diversity of authorial backgrounds at the time.

Another key issue in corpus compilation is the length of text extracts. It is well known that representativeness is dependent on the research question, and as a rule of thumb, high-frequency features can be studied with shorter extracts, while longer extracts are needed if one wants to examine discourse features or rarer lexical items. Following the model set by many previous historical corpora – e.g. the Helsinki Corpus (Kytö 1993), MEMT and EMEMT – we have included short texts in their entirety and taken 10,000-word extracts of longer texts, typically from the beginning. Even though this needs to be remembered especially by those corpus users who are interested in text structure, our chosen cut-off point makes the corpus a reliable source for the study of common grammatical and discursive features, and it is also sufficient for the study of mid-frequency lexical items (see Biber 1993).
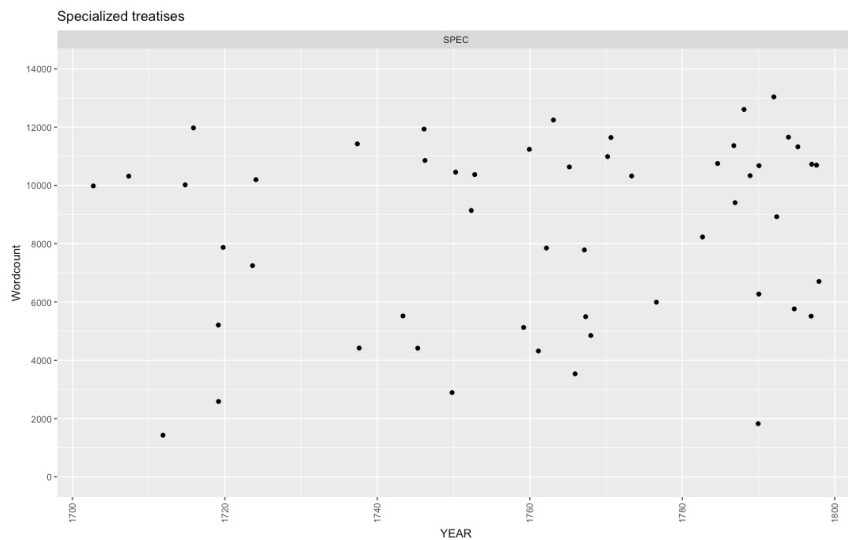
Figure 2 shows the distribution of texts in the category GENERAL TREATISES AND TEXTBOOKS. Each point in the plot represents one text, and the position of the point on the vertical axis indicates the length of the texts. As can be seen, the majority of text contain approximately 10,000 words, which was the cut-off point we used when sampling book-length texts, but a considerable number of shorter texts are also included.
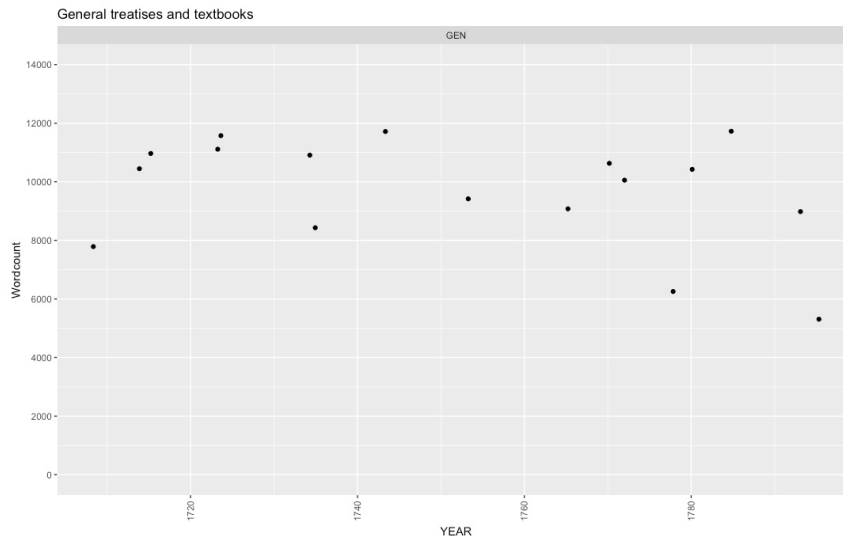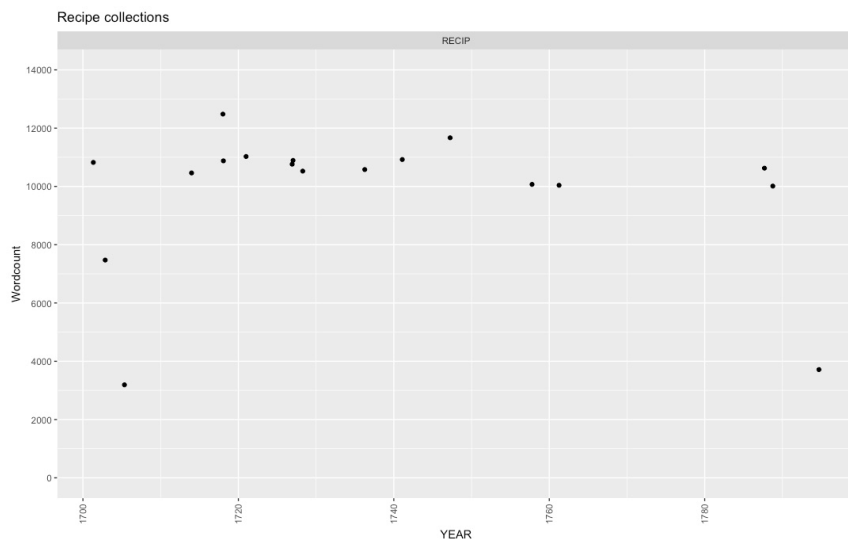
Figure 2. Diachronic distributions and word counts: General treatises and textbooks

Figure 3 shows the chronological distribution of texts in the category SPECIALIZED TREATISES. The plot includes all four subcategories (diseases, methods, therapeutic substances and midwifery), resulting in a larger number of texts than the previous category. This category, too, includes texts of different length.
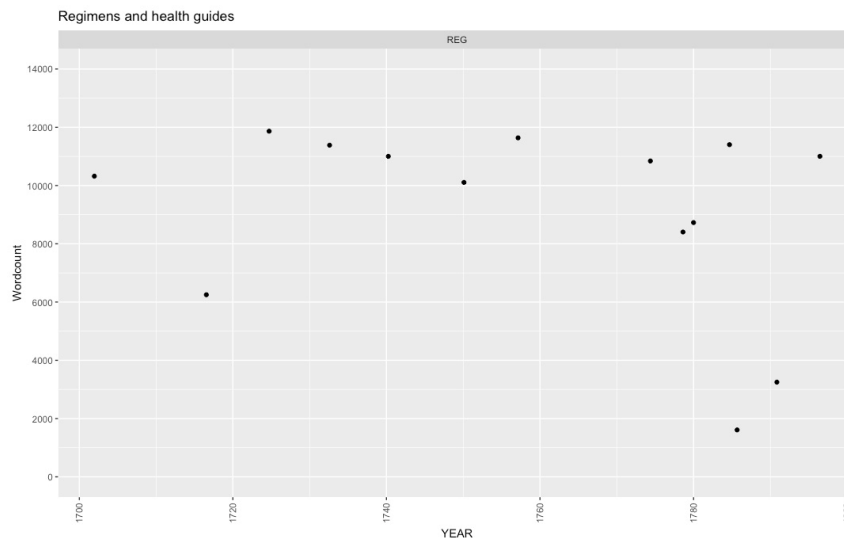
Figure 3. Diachronic distributions and word counts: Specialized treatises

The category RECIPE COLLECTIONS is represented in Figure 4. The figure shows that most texts in this category are extracts of book-length collections, and the distribution is fairly even except the last quarter of the century.
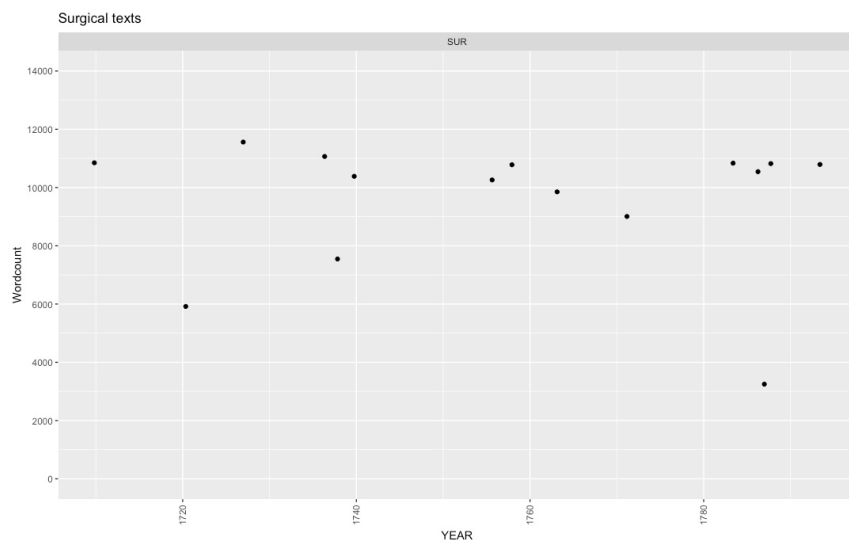

Figure 4. Diachronic distributions and word counts: Recipe collections

The distribution of texts is relatively even in REGIMEN TEXTS AND HEALTH GUIDES, as can be seen in Figure 5.
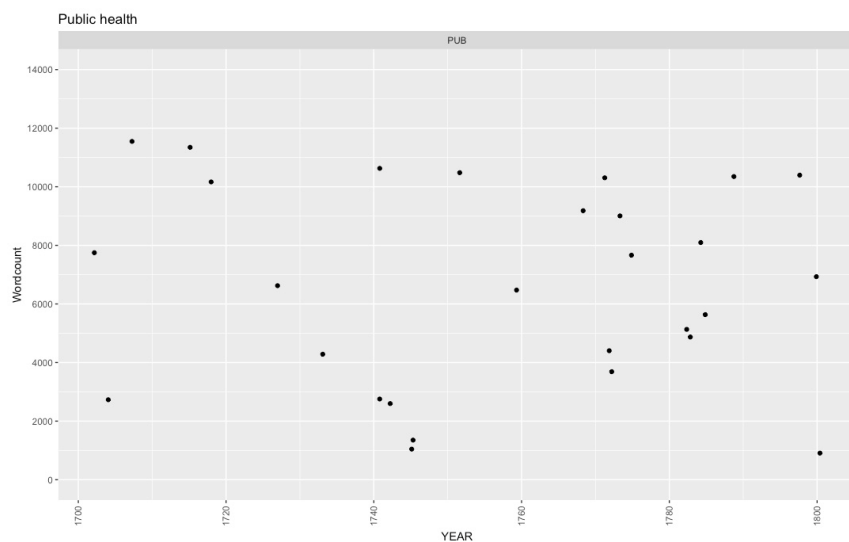

Figure 5. Diachronic distributions and word counts: Regimens and health guides

Similar distributions can be observed in REGIMEN TEXTS AND HEALTH GUIDES (Figure 5), SURGICAL TEXTS (Figure 6), whereas more variation in text length can be observed in PUBLIC HEALTH (Figure 7).
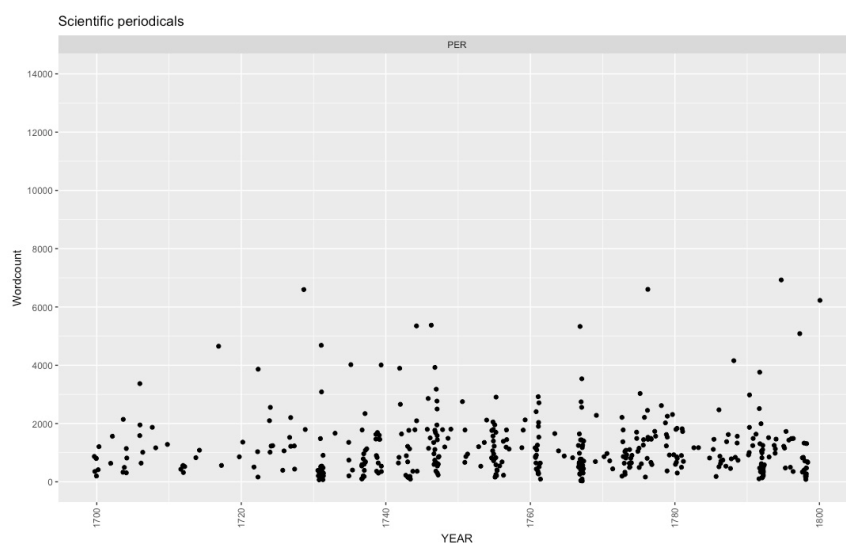
Figure 6. Diachronic distributions and word counts: Surgical treatises

Figure 7. Diachronic distributions and word counts: Public health

Finally, as can be seen in Figure 8, PERIODICALS is different from the other categories in two major ways: there are many more texts, which are much shorter on average, as can be expected.

Figure 8. Diachronic distributions and word counts: Scientific periodicals[3]

The annotation and mark-up is described in a separate chapter. Users planning to carry out diachronic studies using LMEMT in tandem with the two previous corpora are recommended to consult the descriptions of MEMT and EMEMT as well as previous studies based on these corpora (see Taavitsainen and Pahta 2004 and 2011 and the Corpus Resource Database CoRD).

4. The accompanying book

Diachronic developments of medical and scientific writing conventions provide a rich and fascinating object of study. So far, eighteenth-century medical writing has received less scholarly attention compared to earlier centuries, or indeed later times. This state of affairs is in part due to a lack of a large systematically collected database

---

[3] The selection of texts from *The Gentleman's Magazine* is different and does not aim at representativeness.

and, consequently, the history of the special language of medicine in this century is still fairly uncharted. The research reported on in this volume will give precision to some old observations, using data sets that are larger and more systematically collected than in most earlier studies, and we also draw attention to aspects of language use that have not been considered in previous studies.

The book *Late Modern English Medical Texts: Writing Medicine in the Eighteenth century* is an interdisciplinary volume: along with historical linguists and pragmaticians of the research group, the contributors include experts in medical history, computer science, and digital humanities. This reflects our view of the importance of medical history in the study of medical writing, as well as the benefits of collaboration across disciplinary boundaries. The volume focuses on the eighteenth century, but it also takes into account, to some degree, previous periods covered by MEMT and EMEMT. Studies make use of computer techniques in different ways, utilising different retrieval tools and quoting supporting evidence. The scope is wide from corpus-based but mainly qualitative studies to assessments relying on advanced statistical and computational methods. The range extends from lexico-grammatical features and collocations to semantic, pragmatic, and sociolinguistic aspects of language use, unfolding discourse and representations of attitudes and underlying ideology, as shown in the articles of this volume. These studies take the multilayered context of writing into account: the narrow linguistic cotext with what precedes and what comes after, the discourse context (paying attention e.g. to whether the language feature under assessment is found in embedded narrative, in a dialogue or in indirect speech); genre as a functional category (such as recipes of textbooks), and text type (instruction, narration, exposition, description, and argumentation). The period, however, provides the overarching context (see Chapter 2 in this volume).

4. Structure of the book

This chapter has discussed corpus compilation principles and given information about the structure of the LMEMT corpus. The next chapter **Medical writings at the dawn of the new century in context** by Irma Taavitsainen, Peter Murray Jones, and Turo Hiltunen shifts the focus to the cultural context of medicine and medical writing in the eighteenth century. The period is considered from the historical point of view, contextualized with respect to with its sociohistorical and sociopragmatic

backgrounds. According to our approach, a key for interpreting language use is provided by the context, which in turn includes factors like authors and audiences, situations, disciplinary developments, and linguistic cotext and functions.

Continuity and change are at the core in the third chapter that deals with **Topics of medical writing in the eighteenth century**, written in an interdisciplinary frame in collaboration between computer science, medical history, and historical discourse analysis by Irma Taavitsainen, Gerold Schneider, and Peter Murray Jones. The analysis opens with an overview of previous accounts of eighteenth-century developments both within society and in medicine. These earlier views are partly corroborated and partly challenged by a comprehensive empirical study on corpus data with Topic Modeling; this is the first application of this corpus-driven method to historical medical writing. The most important discovery is the strength of the trend towards professionalization in this century and its multiple manifestations. Gradually these developments lead to modern approaches to medicine, starting already in the first half of the eighteenth century by probability counts of the effects of smallpox inoculation. Another striking finding is the continuity of humoral medicine, which remains strong although other trends emerge by its side.

In **Household medicine and recipe culture in eighteenth-century Britain**, historian Alun Withey considers the complex social, cultural, and intellectual contexts of recipe writing. Remedies were eagerly collected as part of a wider interest in welfare; rising prosperity in the era of vibrant economy provided increasing opportunities for expanding intellectual boundaries. New physical theories were presented in medical literature that sought to edify both lay and professional readers. The consumption of published remedy texts was never massive, but served both to promote new ideas and to provide a source for medical knowledge. Newspaper advertising praised proprietary medicines and brought metropolitan products to people in towns and villages across the country. Apothecary shops had a vast range of oils, syrups, pastes, pills, and medicaments for sale and reduced the need for household preparations. The impact of incipient steel industry was also felt as new technologies prompted people to attempt to correct or transform their own bodily forms by new, "enlightened" materials. Such new equipment belongs to innovations, but there was also much continuity, and medical knowledge remained firmly rooted in humoral

medicine. Home remained the central place of medical care throughout the eighteenth century, but little by little public medicine began to make an impact slowly, and then mostly in urban areas.

Medical case reports in Late Modern English were another well-established genre with a long diachrony in the vernacular, but its functions and linguistic realizations vary in different periods; they are dealt with in this chapter by Anu Lehto and Irma Taavitsainen. Instruction continued as the most common purpose of writing, but new functions emerge as well, as case reports are used to demonstrate new methods of treatment and as efficacy proofs of novel cures. Attention is paid to the degree of conventionalization and the perspective through which the narrative is told, and in many texts the focus shifts to the patients. By combining quantitative corpus linguistic techniques and qualitative discursive analysis, the study shows that a transition took place from the earlier thought styles to more modern approaches, as there is an increase in numerical assessments towards the end of the century. The developments are, however, somewhat different in various layers and fields of medical writing, and the linguistic form of case studies varies considerably between text categories. Importantly, patients begin to record their own experiences and for the first time we have "ego-documents" of the kind.

The following chapter, **Regimens and their readers in eighteenth-century England** by David Gentilcore illuminates the history of health guides. Health guides were already a well-established genre of medical writing in the eighteenth century, although their popularity had declined in the previous century. The eighteenth century saw a revival of regimen texts mainly targeted at the upper and upper middling classes of society, and genre contents changed to reflect new trends of medical thinking with chemical interpretations of foods and their influences. This chapter focuses on two of the most influential regimen writers of this century, George Cheyne and William Buchan, whose books offered advice on diet and lifestyle. The former wrote a bestseller in 1724, while the latter also gave recommendations to improve the diet of the poor about half a century later, thus opening dietary advice to the reach of the English nation as a whole; this marks a new attitude to preventative medicine. However, regimen texts mainly provided reading for more affluent audiences who had the means and leisure to follow their advice.

An account of **Polite society language practices: Letters to the Editor on medical issues in *The Gentleman's Magazine* 1731-1800** discusses correspondence between readers and editors of the new publication channel, *The Gentleman's Magazine* (GM), where issues of health were debated in the written form on a broad front. This pragmatic study by Irma Taavitsainen applies an ethnographic and socio-constructivist approach to politeness as a discursive practice, taking people's own notions of what was appropriate and desirable for smooth interaction in polite society as its point of departure. The method of analysis is qualitative corpus-based discourse analysis. Politeness, sociability and concern for public good were values underpinning eighteenth-century gentility culture where it was important to recognize one's own position and act accordingly. The diseases that people wrote about were mostly minor discomforts and everyday nuisances, but dietary advice and first aid tips were also given to "fellow-sufferers" (to quote a pseudonym). Polite speech acts like compliments and thanks prevail, but impolite speech acts occur, too, in debates and disagreement, but they are mitigated or veiled in politeness. These letters give us a glimpse into mindsets and we have direct access to opinions on what GM readers considered worth attention.

An application of a corpus-linguistic method by Anu Lehto is presented in the following chapter. She assesses three-word lexical bundles in order to detect what they can reveal about **Changing portrayals of medicine and patients in eighteenth-century medical writing**. Her material includes texts in PUBLIC HEALTH and METHODS, as well as extracts of case studies in other categories. The approach is corpus-driven, i.e. it is conducted without predefined categories according to what the corpus yields. Three main functional categories emerge: referential and textual bundles, and stance expressions, and there is variation according to the different corpus categories. PUBLIC HEALTH texts discuss medical matters on the social level, while METHODS concentrate on patients. Likewise, case studies focus on the patient, but in a narrative mode. The assessment reveals some novel practices as increasing importance is attached to observation, especially in METHODS and case studies; statistical methods and quantification are enforced by repeated constructions, and bundles in the novel category of PUBLIC HEALTH refer to the current hospital movement and issues of hygiene.

Stylistic features in texts for different audiences in various publication channels are discussed in a chapter called **Professional and lay medical texts in the eighteenth century: A linguistic stylistic assessment** by Irma Taavitsainen. Inoculation, longevity, sea-bathing, water, and air were fashionable topics and occur in various types of writing across the corpus. Authors of these texts were mostly educated professionals who wrote for their peers, but household texts for general audiences and literate women are also included in the corpus. Monographs provided the most important channel of publishing medical texts, but from the 1730s onwards periodicals, both specialized and general, had an increasing role in medical communication. The aim of the study is to explore differences between medical writings for professional and lay audiences. The method of assessment is mainly corpus-aided qualitative discourse analysis, and besides sociolinguistic variation, attention is paid to diachronic developments. The results show that changes take place in professional writing, while texts for general audiences are apt to retain the old practices.

The final chapter **The *symptom* comes of age: Sign semantics from the Late Medieval period to the Late Modern** by Jukka Tyrkkö provides a diachronic study on the repertoire of medical signifier terms, a lexical field has undergone several changes in the course of time. Based on the data of the three corpora – MEMT, EMEMT, and LMEMT – the study traces the history of signifier terms from Middle English period to the end of the Late Modern period both with a quantitative corpus linguistic assessment and qualitative discourse analysis. The data displays both stability and change: while *sign* and *accident* were used throughout the 400-year time period, the term *symptom* appeared as a new term in the 1600s and gained more ground in the eighteenth century. These changes reflect both the more theoretical side of the discipline and its practical applications.

We strongly agree with Marc Alexander's (2018) view that what distinguishes a great corpus from a good one is the quality of documentation, and therefore the second part of the book is devoted to the description of the *Late Modern English Medical Texts* (LMEMT). It provides detailed descriptions of the individual text categories by authors primarily responsible for them. The descriptions follow a set pattern, covering text selection and representativeness and providing information about authors,

audiences, and reception. The volume concludes with a description of the annotation and mark-up conventions by Turo Hiltunen and Jukka Tyrkkö, which also contains suggestions for using the corpus in different research scenarios.

5. Conclusion

LMEMT is a new resource covering a broad range of medical texts from cutting-edge scientific treatises to everyday household texts. Our approach to medical writing always takes into account the context as a multifaceted phenomenon (see Chapter 2 in this volume). Sociolinguistic parameters provide important facts for anchoring texts to their users, discourse communities, as well as authors and audiences in various constellations of textual practices. For us, the main contribution that we wish to make with the release of LMEMT to the research community is convenient access to a representative corpus containing materials that have previously received little scholarly attention. This corpus is large enough to enable the analysis of language, style, and discourse over the one-hundred-year period in focus. Besides finding answers to new linguistic research questions, the corpus also enables interdisciplinary research at the interface between corpus linguistics, computer science, philology, history of science, book history, and digital humanities.