Early Modern English Relativisers:
A Corpus–Based Study on *which* and *the which*

Noémi Berti

Master's Thesis in English Linguistics

Department of Foreign Languages

University of Bergen

May 2024

Abstract in English

The aim of this thesis is to investigate the development of relative *the which* in Early Modern English scientific prose. The relative pronoun was common in Late Middle English and throughout the 1500s, but is reported to have disappeared during the 17th century (see, e.g., Rissanen 1999: 297). Past studies (see, e.g., Curme 1912; Reuter 1937; Mustanoja 1960; Rydén 1966 or Raumolin–Brunberg 2000) have only reached limited conclusions on the relativiser's development or uses as opposed to plain *which*. Further, past accounts often contradict each other.

The highly specific context of Early Modern English scientific writing was expected to contain more instances of *the which* than lower registers. The data was collected from the Málaga Corpus of Early Modern English Scientific Prose (MCEModESP) (Calle–Martín et al. 2020) and the Early Modern Multiloquent Authors (EMMA) Corpus (Petré et al. 2018). The first dataset allowed for a quantitative approach, while in the second the methods were mostly qualitative. Possibly due to its conservative nature (Taavitsainen 2000: 147), the selected genre offered important insights into, especially, later uses of the relativiser.

This project contributes to the discussion on the relative pronoun by providing two significant findings. The analysis revealed that *the which* may have been in use longer than previously assumed, namely until the end of the 17th century. Further, several instances in the data illustrated fine differences between *the which* and plain *which*, so far often treated as free variants. These differences can be traced back to the information parameter and indicate a preference for *the which* in non–restrictive clauses, succeeding clauses containing *which* (or *that*) when referring to the same antecedent.

Keywords: *the which*; relative pronouns; Early Modern English; Early Modern English scientific writing; historical sociolinguistics

Abstract in German

Diese These untersucht die Entwicklung des Relativpronomen *the which* in frühneuenglischer Wissenschaftsprosa. Das Relativpronomen war im Spätmittelenglischen und während des 16. Jahrhunderts üblich, verschwand der Literatur zufolge jedoch im 17. Jahrhundert (siehe z. B. Rissanen 1999: 297). Frühere Studien (siehe z. B. Curme 1912; Reuter 1937; Mustanoja 1960; Rydén 1966 oder Raumolin–Brunberg 2000) kamen nur zu begrenzten Schlussfolgerungen über die Entwicklung oder den Gebrauch des Pronomens, besonders im Gegensatz zum einfachen *which*. Darüber hinaus widersprechen sich die bisherigen Aussagen häufig.

Es war zu erwarten, dass der hochspezifische Kontext der frühneuenglischen wissenschaftlichen Literatur mehr Instanzen von *the which* enthalten würde als niedrigere Register. Das Datenmaterial stammt aus dem Málaga Corpus of Early Modern English Scientific Prose (MCEModESP) (Calle–Martín et al. 2020) und dem Early Modern Multiloquent Authors (EMMA) Corpus (Petré et al. 2018). Der erste Datensatz ermöglichte einen quantitativen Ansatz, während die Methoden im Zweiten hauptsächlich qualitativ waren. Womöglich aufgrund seiner konservativen Eigenschaften (Taavitsainen 2000: 147) bot das gewählte Genre wertvolle Einblicke in, insbesondere, die spätere Verwendung des Pronomens.

Dieses Projekt trägt zur Debatte über das Relativpronomen bei, indem es zwei bedeutende Erkenntnisse liefert. Die Analyse ergab, dass *the which* möglicherweise länger in Gebrauch war als bisher angenommen, nämlich bis zum Ende des 17. Jahrhunderts. Außerdem wurden in beiden Korpora mehrmals feine Unterschiede zwischen *the which* und dem einfachen *which*, die bisher oft als freie Varianten behandelt wurden, festgestellt. Die Unterschiede lassen sich auf die Restriktivität der Relativsätze zurückführen und weisen auf eine Bevorzugung von *the which* in explikativen Relativsätzen hin, die auf Sätze folgen, die *which* (oder *that*) enthalten, wenn sie dasselbe Bezugselement beschreiben.

Schlüsselwörter: *the which*; Relativpronomen; Frühneuenglisch; Frühneuenglisches wissenschaftliches Schreiben; historische Soziolinguistik

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Jerzy Nykiel, for his support, guidance, and invaluable advice throughout the entire process. Thank you for believing in this project from the beginning, for your encouragement, and for your patience.

I would also like to thank the Sammen team for their help and support during my study period in Bergen. To Berit, Inghild, and Gunnhild especially, for your guidance and for always listening.

And finally, thank you to José, Martin, and Patricia, for your unconditional love and for being there through all highs and lows.

Table of contents

List of tables and figures

This page is intentionally left blank.

1. Introduction

Relativisation is among the best studied topics in the history of the English language (Nevalainen & Raumolin–Brunberg 2002: 109). Within the context of Early Modern English, several large–scale studies have successfully traced the development of the main relative pronouns (see, e.g., Saito 1961; Rydén 1966 or Dekeyser 1984). The special focus on this period is not surprising – the years between 1500 and 1700 saw the emergence of a standard variety, and the relative pronoun system reached the stage it largely maintains to this day (see, e.g., Fennell 2001: 138; see also Dekeyser 1984: 61).

The English vernacular gradually replaced Latin as the language of administration and education. This, in turn, resulted in a desire to regularise the language. Alongside important cultural events, such as the blooming of the printing industry, English was further influenced by the strongly perceived politeness scale (see, e.g., Rissanen 1999: 294; Fennell 2001: 138 or van Gelderen 2006: 155–6; 180). Borrowing from Latin and French models, which were perceived as more suitable for formal registers, the relative pronoun system increasingly parted with that governing Middle English (Görlach 1991: 95; 124–5).

While the introduction of the animacy distinction and the rise of relative *who* are well–documented (see, e.g., Nevalainen & Raumolin–Brunberg 2002), a development that has often been overlooked is the gradual disappearance of the relativiser *the which*. Only one in–depth article, namely Raumolin–Brunberg's (2000) corpus–based study, has been published about the pronoun, that I am aware of. Numerous questions remain unanswered, including the exact origins of the relativiser, its decline after 1600, and its uses as opposed to plain *which*.

The aim of this thesis is to contribute to the discussion on the development of relative *the which*, testing past accounts and employing improved methods. Special attention was paid to the contexts where the relativiser is mostly expected according to the literature, namely higher registers and academic prose, including scientific writing (Fischer 1992: 303–4). Keeping this in mind, I collected and analysed data from The Málaga Corpus of Early Modern English Scientific Prose (MCEModESP) (Calle–Martín et al. 2020) and the Early Modern Multiloquent Authors (EMMA) Corpus (Petré et al. 2018). In the case of the first dataset, my approach was quantitative. With the second corpus, on the other hand, it was mostly qualitative. The following research questions have guided my study:

1. How are the relativisers *which* and *the which* used in Early Modern English scientific/medical prose and to what extent do they act as free variants? How are they distributed in terms of antecedents and are there any patterns? If yes, which ones confirm previous conclusions and which ones do not?

2. How do the distribution of *the which* and its gradual disappearance match the timeline observed in other studies?

3. How are the changes in the Early Modern English relative pronoun system visible in individual authors' writing, and what role does their socio–economic background play?

Basing myself on the so far available literature on the topic, I propose the following hypotheses as an attempt to answer the research questions:

Hypothesis 1:

a. *Which* and *the which* are used for both personal and nonpersonal antecedents until they are replaced by *who* with human referents, a change which becomes visible particularly in data from the end of the period.

b. *Which* is far more frequent in the data and preferred as the subject in relative clauses. *The which*, on the other hand, is rarer and mostly used in non–restrictive continuative clauses and/or prepositional phrases. Apart from these trends, it would seem like the relativisers do, at least in theory, act as free variants.

c. Due to its rather formal style and specialised vocabulary, scientific/medical prose might be more conservative than other genres and present less variation. Simultaneously, it might be influenced earlier by standardisation processes and early prescriptive conventions, as these texts are officially published, directed at a wider audience, and aim at a certain level of prestige.

Hypothesis 2:

a. The period from 1500 to 1600 is expected to present the most variation between the relativisers and, during the following decades, their use gradually becomes more regularised. This also applies to the animacy distinction and the related choice between (*the*) *which* and *who*.

b. The development of the pronouns, including the rapid disappearance of *the which* after 1600, is expected to match the general timeline observed in previous studies. However, given the

formal and perhaps more conservative nature of scientific writing, the latter might be delayed compared to other genres.

Hypothesis 3:

a. *The which* is expected to become increasingly rarer in individual authors' writing and to, eventually, disappear. Only very few sporadic occurrences, if any, are expected towards the end of the 17th century.

b. Over time, the authors are expected to use (*the*) *which* less with personal antecedents and to gradually transfer this function increasingly, if not exclusively, to *who*.

c. Given the high social rank of the authors and their educational background, they are generally expected to comply early with the conventions suggested by the 17th– century grammarians.

In the next chapter, I will provide an outline of the Early Modern English relative system. I will also discuss the historical context and main linguistic developments that characterise Early Modern England. Further, I will introduce the most important relative pronouns for this study, including *the which*. The focus of section 2.2 will be Early Modern English medical writing and its relevance for historical sociolinguistics. Finally, in section 2.3, I will introduce the concept and field of linguistic lifespan change, relevant for the second half of this project.

In chapter 3, the focus will shift to the methodology employed in this study. I will introduce the two corpora mentioned above, and in section 3.2 I will also discuss the relevance of statistical analyses in this project.

Lastly, chapter 4 is dedicated to the display and discussion of my results, which will be illustrated with the help of tables and examples. Furthermore, in section 4.3 I will address the statistical significance of the results and additionally test their relevance. I will conclude the thesis with a summary of my main findings in chapter 5.

2. Theory

In this chapter, I will provide an outline of the linguistic situation in Early Modern England. First, I will discuss the context of a period characterised by processes that, eventually, led to a standard variety. In sections 2.1.1 to 2.1.5, the focus will shift to the relative pronoun system and the special case of *the which*. Subsequently, I will introduce the genre of Early Modern English medical writing and its properties, highlighting its development and impact on society. Finally, section 2.3 focuses on the field of linguistic lifespan change, with special regards to individuals living, speaking, and writing between 1500 and 1700.

2.1 Early Modern English: towards a standard variety

The Early Modern period (c.1500–1700) was a time of major social changes. The beginning of the period is characterised by migration to the larger cities, that would grow exponentially and thrive economically in the following two centuries. The increased mobility, in turn, resulted in looser ties among citizens and more contact between diverse social and dialectal backgrounds (see, e.g., van Gelderen 2006: 155).

On the cultural level, the Early Modern period was the time of the Renaissance and saw changes in the world view, as well as a revival of classical Greek and Latin texts. Immense advances in education, vernacular literature, and entertainment, such as theatre, were the result. Furthermore, the years around 1600 were the time of William Shakespeare (1564–1616) who, to this day, is celebrated as perhaps the main representative of English literature. In 1611, the King James Version of the Bible was published. Written in English, it today represents an invaluable account of early vernacular prose and standardisation of the language (Fennell 2001: 136–7; 147–8; see also van Gelderen 2006: 155–8).

From a linguistic perspective, the most important changes that took place include the completion of the Great Vowel Shift, changes in the pronoun system, and the rise of a written standard variety (see, e.g., Görlach 1991; van Gelderen 2006 or Nevalainen 2006). According to Leith, a standard language can be defined as a variety of a given language that has minimal variation of form and maximal variation in function (1983: 32; cited in Nevalainen & Raumolin–Brunberg 1989: 82). In other words, this means that a language community uses a relatively uniform variety as the primary language in all contexts, both locally and nationwide. Until the 1650s, English did not have such a standard. Instead, the principal language used in

official contexts, including science, education, as well as law and administration, was Latin. To a lesser extent, French was also the preferred choice over the vernacular (van Gelderen 2006: 180; Nevalainen 2006: 29). Without an English standard variety, a considerable amount of free linguistic variation was the rule rather than the exception, especially at the beginning of the period. By 1700, however, English had replaced Latin as the former standard language in official contexts and the spelling system was very close to that of Present–day English (see, e.g., Fennell 2001: 138).

Early Modern England also saw important scientific advances and numerous new inventions; the perhaps most impactful one being the printing press, which made texts available to a wider audience. Introduced by Caxton in 1476, publishing truly took off in the 1640s and the number of titles distributed by publishing houses steadily increased by the decade (van Gelderen 2006: 155–6; Pahta & Taavitsainen 2011: 4; Taavitsainen et al. 2011: 14). Unsurprisingly, printing and publishing are also among the biggest catalysts for the regularisation of spelling, as publishing houses needed spelling conformity.

The attempts at establishing standard spelling conventions were numerous: word lists and works on "correct" spelling became popular in the second half of the 16th century, eventually culminating in prescriptive grammar books and dictionaries in the following centuries. The early grammars, however, were more descriptive than prescriptive (van Gelderen 2006: 180–4).

The impact of standardisation processes becomes particularly evident when comparing texts from the beginning and the end of the period. While later texts are mostly in line with the Modern English spelling system, 16th century writing is still characterised by extensive variation and constructions inherited from Middle English (Rissanen 1999: 187; see also Fennell 2001: 138). It is important to highlight that the emerging standard only concerned the written language. Informal spoken English has retained a richness of variants to this day (Rissanen 1999: 188).

On the societal level, the gradual establishment of a standard variety had both advantages and drawbacks. On the one hand, a uniform language governing all official sectors meant increasing accessibility and stability. However, the implementation of a standard spelling happened at the expense of regional variation. It, further, created even greater inequality between those who could use the standard variety and those who, on the other hand, could not. Using the standard became increasingly associated with a higher social rank and status, leaving

the part of the population that had limited access to education and/or contexts where to use it at a clear disadvantage in the eye of society (Fennell 2001: 154; see also van Gelderen 2006).

The year 1700 traditionally marks the end of the Early Modern English period, as the Great Vowel Shift was mostly complete by the turn of the century and the spelling system was relatively uniform (van Gelderen 2006: 155).

2.1.1 Relative clauses

As illustrated in the previous section, the Early Modern period is a time of important linguistic changes, eventually leading to a standard variety. The area this study is particularly concerned with is that of relativisers and their diachronic development. Before looking into the individual pronouns, it is necessary to outline the different relative clause types and their properties.

A relative clause is a subordinate clause that functions as a modifier to an antecedent (Huddleston & Pullum 2005: 306). According to Fischer, in Present–day English they are formed according to two parameters: the animacy parameter and the information parameter. The former determines the choice of the relative pronoun (i.e., *who* and its inflected forms against *that* and *which*; see section 2.1.3.3). The latter, on the other hand, concerns the restrictiveness of the clause. So–called restrictive (or defining, see the *Oxford Dictionary of English Grammar*) clauses characterise their antecedent. The information they provide is essential to identify their referent. Non–restrictive (or non–defining) clauses, on the other hand, merely add further information about an identified antecedent (1992: 295–6; see also Quirk et al. 1985: 1239–42). Examples for the two clause types are illustrated in sentences [1] and [2]. The relative clause in [1] is restrictive, while in [2] it is non–restrictive. Both clauses are marked in italics:

[1] Can I borrow the book *that you bought yesterday?*

[2] This book, *which is very old*, belongs to my sister.

In example [1], the information contained in the relative clause is essential to identify the book in question. Without it, it would be hardly possible for the recipient to understand which book is meant (of course, in a real–life situation this would also depend on the context). In example

[2], on the other hand, there is no ambiguity, and the relative clause merely adds information about the referent. However, it is clear which book is meant, even without the clause.

The syntax of English relative clauses is heavily influenced by the Latin model. The Early Modern English period saw an increase in other finite and non–finite constructions. However, especially non–restrictive relative clauses remained the most important type of subordination (Görlach 1991: 124–5).

At the beginning of the period, the language still presented high variation among the two parameters. The animacy distinction, influencing the pronoun choice, spread during the 17th century (see, e.g., Saito 1961: 84; cited in Dekeyser 1984: 71; or Nevalainen 2006: 84–5). Furthermore, the convention of separating a non–restrictive clause with a comma is a later convention and still uncommon in 16th–century texts (Nevalainen 2006: 84). It is, therefore, not surprising that early examples may appear ambiguous to the modern reader.

2.1.2 Antecedents in relative clauses

As described in section 2.1.1, the role of relative clauses is to modify their antecedent (Huddleston & Pullum 2005: 306). In restrictive clauses, the contained information is essential to identify the referent, while in non–restrictive clauses it is not.

In a relative construction, the antecedent is the element the relative clause refers back to (Hasselgård et al. 2012: 320). However, as stated by Quirk et al., "the relationship […] between a pronoun and its antecedent […] is not one which can be explained by the simple act of replacement". With the context being key, pronouns can be defined as devices for recapitulating and referring to neighbouring clauses, reducing repetition and grammatical complexity (1985: 76). Also, in most cases the antecedent precedes the pronoun, but under certain conditions it can follow it, too (Huddleston & Pullum 2005: 101).

Antecedents are usually noun phrases (including also human referents and pronouns), but they can also be sentential (see, e.g., Stirling & Huddleston 2002: 1458–9 or Hasselgård et al. 2012: 272–3; 320; 324). Example [3] shows a relative construction with a noun phrase as its antecedent (underlined). The following example [4], on the other hand, has an entire clause as its referent. In Early Modern English, instances like these are typical of higher registers, such as academic or scientific prose (Fischer 1992: 303–4):

[3] And if this ointment will not heal the canker then strew thereon a powder *which* is good to fret away cankers. (Unknown author, between 1499 and 1599. *MS Hunter 135*. MCEModESP)

[4] […] other physicians judged him not possible to live one year *which* he did confess a little before his death […] (Bartholomew Blogate, between 1599 and 1699. *MS Hunter 64*. MCEModESP)

2.1.3 Relative pronouns in Early Modern English

The development of Early Modern English is, generally, characterised by a desire to regularise and systematise the language. Special importance is attributed to extralinguistic factors such as politeness and conventions of formality (Rissanen 1999: 294). These aspects are largely motivated by English replacing Latin as the primary language in official contexts. As a result of this switch, there was a need to adapt and tackle the previously perceived "inadequacy" of the vernacular (Görlach 1991: 95).

Particularly interesting for this study is the development of the pronoun system. The fact that the Early Modern period saw changes in this domain is interesting in itself, given that pronouns are grammatical (closed–class) words. While this means that their number cannot be freely increased through borrowing and word–formation processes, it also means that they are less likely to fall into disuse than nouns (Busse 2012: 731–2).

Perhaps the most studied case is the *you*/*thou* paradigm. In Old English, *thou* and *thee* were the singular forms for the second person, while *ye* and *you* were plural. Later, in Middle English, the plural pronouns became the polite forms, similar to the French *tu*/*vous* model. However, the distinction was never as rigid as it was in French and disappeared during the Early Modern English period. *Ye* was lost around 1600, and *thou* and *thee* also slowly fell out of use. *You*, on the other hand, eventually become the common form and acquired a more neutral function (van Gelderen 2006: 121–2; 167; see also Sato 2016: 214).

The possessive pronoun inventory also saw an important change, namely the introduction of the neuter pronoun *its*. Emerging in the late 16th century, it most likely came into existence as an analogy of *yours*, *hers*, etc. During the following decades, it gradually replaced *his* when referring to inanimate and/or nonhuman referents (Lass 1999: 148; Görlach 1991: 85–6; both cited in Cowie 2012: 605; van Gelderen 2006: 167).

A different situation, however, is that of the relative pronouns. Just like the relative clause structure itself (see sections 2.1.1 and 2.1.2), the Early Modern English relative pronoun inventory is rather similar to the one we have today (Nevalainen 2006: 84). As noted by Dekeyser (1984: 61), "no new elements have been added to the set of relativisers since the 15th century. The crucial difference between [Middle English] and [Present–day English] is not the number of relatives, but the system that governs their distribution". In other words, no new items were introduced, but the rules conditioning their use underwent a number of important changes.

In comparison to other areas of research, relativisation is one of the best studied topics in the history of the English language (Nevalainen & Raumolin–Brunberg 2002: 109). Thanks to the abundant literature, the profiles of the individual relativisers can be illustrated in detail.

2.1.3.1 The relativiser *that*

The most common relative pronoun in Early Modern English was undoubtedly *that*. Descending from Old English *þæt*, it could be used in both restrictive [5] and non–restrictive [6] relative clauses (see section 2.1.1). Furthermore, it was popular with both human [5] and nonhuman [6] referents. In Middle English, free combinations such as *who that* or *which that* had been possible, and the relativisers could also be deleted under certain conditions (Dekeyser 1984: 61; see also Fischer 1992: 302).

[5] This is good for anyone *that* is troubled with a pain in the side […]. (Unknown author, 16th century. *MS Hunter 95*. MCEModESP)

[6] Also, there is a Neutral sort, *that* has neither Sympathy nor Antipathy, but is Indifferent. (Cavendish, 1668. EMMA)

2.1.3.2 The *wh*–pronouns

Despite *that* generally being the preferred relativiser, especially in higher registers the *wh*–pronouns (e.g., *who*, *whose*, *whom* or (*the*) *which*) were also common (Romaine 1982: 71). The spread of these pronouns was most likely motivated by the heavy functional load of *that*, which may have caused misunderstandings in ambiguous contexts (Rissanen 1999: 295).

An interesting factor is the system that governs their use. The *wh*–pronouns are of Germanic origin; however, the way in which the paradigm has been implemented follows Latin and French models. Especially Latin influence may also have been the driving force behind the use of interrogative pronouns as relativisers. In English, this use is first observed in 12th–century texts (Mustanoja 1960: 192; 195; cited in Nevalainen & Raumolin–Brunberg 2002: 110; Rissanen 1999: 189).

The foreign influence on the relative pronoun system may be traced back to more formal, written registers. When it comes to literary translations, direct loans and borrowing were not uncommon. This, in turn, resulted in a high amount of lexical but also morphological and syntactic borrowing, with foreign uses and patterns being adopted into the language (Nevalainen & Raumolin–Brunberg 2002: 110).

Dekeyser observed in his (1984: 62) quantitative study on Early Modern English relativisers that, in the 16th century, the *wh*–pronouns were well established in all types of non–restrictive relative clauses. *That*, on the other hand, was particularly common in texts representing more colloquial, spoken language (cited in Rissanen 1999: 293).

Taking a closer look at the clause types, Dekeyser also observed that *who* was mainly used in non–restrictive relative clauses. *That*, on the other hand, was preferred in restrictive ones, while (*the*) *which* was 'neutral'. He also, however, pointed out that this distribution seemed to depend on the specific dataset and the text type. In fact, he noticed a difference between the different corpora he used, especially with regards to written versus spoken discourse[1] and the time frame covered (1984: 69–70).

2.1.3.3 The animacy distinction

Another important development concerns the animacy distinction. In Middle English, (*the*) *which* could be used with both human and nonhuman antecedents, before being replaced by *who* when referring to persons. The late introduction of nominative *who* in 15th–century letters presents an interesting case, as its oblique form *whom* and genitive *whose* are already recorded in 12th–century texts (Dekeyser 1984: 61; Nevalainen & Raumolin–Brunberg 2002: 109). The speculations on the introduction of *who* are manifold. On the one hand, the lack of a nominative form clearly formed a gap in the paradigm (Steinki 1932: 30; Rydén 1983: 130;

[1] The data covering spoken discourse was taken from Saito's (1961) corpus of Modern Colloquial English (Dekeyser 1984: 61–2).

both cited in Johansson 2012: 784). On the other hand, it may have been introduced specifically for the semantic purpose of referring exclusively to animate antecedents, as (*the*) *which* and *that* were no longer considered suitable for this role (Ball 1996: 246–7; cited in Nevalainen & Raumolin–Brunberg 2002: 110). A comparable introduction, in this regard, is that of *its* replacing *his* with nonhuman referents (see section 2.1.3).

The development of the animacy distinction spread relatively quickly during the 17th century (see, e.g., Saito 1961: 84; cited in Dekeyser 1984: 71; or Nevalainen 2006: 84–5). Within the context of standardisation and the later prescriptive movements, the first clear division between (*the*) *which* and *who* was made by James Greenwood during the first decade of the 18th century. No earlier grammars had ever proscribed the use of the former with human referents (Bately 1965; cited in Sato 2016: 208). However, Early Modern grammarians were undoubtedly well aware of the ongoing linguistic changes. In his (1674) *Grammatica Linguae Anglicanae*, Wallis observed with regards to the use of *which* in The Lord's Prayer, that *who* was a more proper and elegant choice (Sato 2016: 208).

Nevertheless, as illustrated in example [7], occurrences of (*the*) *which* with human antecedents could still be found in the 17th century. In Dekeyer's data, the dehumanisation of (*the*) *which* is clearly reflected, as its occurrences with persons rapidly decline after 1560. Interestingly enough, nonhuman *who* also appears throughout his data, but its use is mostly justified by personification, often referring to animal antecedents (1984: 70). *Who* seems to first have replaced (*the*) *which* in relative clauses with proper–name antecedents or those referring to the Deity, as underlined in example [8]:

[7] An oyle for a bruse in the eye a proued by me of <u>a womas</u> *which* had lost the use of her eye. (Townshend Family, 1636. *MS Wellcome 774*. MCEModESP)

[8] And this have I written in the praise of <u>almighty god</u> *which* gave me grace so to devise for my health and prolonging my life. (Unknown author, between 1499 and 1599. *MS Hunter 135*. MCEModESP)

The introduction of the animacy distinction is in line with the general desire to systematise the use of various grammatical forms in Early Modern English. Simultaneously, it also presents parallels to the polite and formal expression of Tudor and Stuart society, where the 'personality' of the referent or addressee played an increasingly important role. The development reached its present–day state in the 18th century (Rissanen 1999: 294), but the data from Nevalainen &

Raumolin–Brunberg's (2002: 112) study on the rise of relative *who* suggested that personal *which* persisted longer (and is, perhaps, still in use) in some regional varieties.

2.1.3.4 The *zero* relativiser

Lastly, relative clauses could also be formed with *zero*, i.e., without an expressed relative pronoun that links the main and subordinate clauses. Confined to restrictive relative clauses, *zero* (marked with *Ø*) occurs in both subject [9] and direct object [10] position, or with a stranded preposition:

[9] Heere they come *Ø* will tell you more. (William Shakespeare, 1623. *All's Well that Ends Well* III.ii. Example taken from Rissanen 1999: 298)[2]

[10] Mr. Conyers. My Lord, the first Meeting *Ø* Mr. Porter speaks of, where the Prisoner Mr. Rookwood was, is at the Globe-Tavern, where this Discourse was; […] (Ambrose Rookwood, 1696. A Corpus of English Dialogues 1560–1760 (CED). Example taken from Johansson 2012: 788)

In the 18th century, grammarians seemed to define the *zero* link in non–subject position as a colloquialism. However, in subject position and, generally, throughout the Early Modern period it also appeared in formal writing (Rissanen 1999: 298–9). The profiles of the Early Modern English relative pronouns are summarised in Table 2.1:

reference	**subjective case**	**objective case**	**possessive case**	**determiner**
human	*(the) which→who* *that* *zero*	*whom* *that* *zero*	*whose*	*(the) which*
nonhuman	*(the) which* *that* *zero*	*(the which) which* *that* *zero*	*whose*	*(the) which*

Table 2.1: The Early Modern English relative pronoun system (incomplete) with regards to the animacy distinction. Taken and adapted from Nevalainen's Table 6.2 (2006: 84)

[2] Instances where the antecedent noun phrase is separated from the zero–introduced relative clause, such as this one, are generally rare (Rissanen 1999: 298).

Table 1 emphasises the establishment of the animacy distinction and illustrates the switch from (*the*) *which* to *who* with human antecedents. Furthermore, the table shows how neither *that* nor (*the*) *which* have a possessive form. Consequently, *whose* becomes the only option in this function, with both human and nonhuman referents (Dekeyser 1984: 61; Rissanen 1999: 294). Lastly, *which* and *the which* are the only pronouns that occur as determiners[3].

2.1.4 The relativiser *the which*

In previous studies, the relative pronoun *the which* has often been treated as a free alternative to *which*, and a distinction between the two has rarely been made. Dekeyser's (1984) quantitative study includes a category labelled "(the) which", but no distinction is made between the two variants. Similarly, Nevalainen & Raumolin–Brunberg (2002) treat them as one and the same variable in their study on the rise of relative *who*. *The which* is, in fact, not explicitly mentioned throughout their analyses. It seems, thus, like the relativiser is underrepresented in the literature. This becomes especially evident with regards to the, otherwise, vast number of books and articles on the diachronic development of the English pronoun system. Other topics, such as the distinction between human and nonhuman referents, or the related establishment of the possessive pronoun *its*, have been studied in far more detail (see, e.g., Nevalainen & Raumolin–Brunberg 1994; 2002 or Sato 2016; 2019; see also sections 2.1.3 and 2.1.3.3).

A potential reason for the relativiser's tendency to be overlooked in studies may be its close resemblance to *which*, being similar both in form and in function. In fact, the contexts in which the two variants were used appear almost identical. The main difference, as observed by Raumolin–Brunberg in her (2000) study, seems to be a trend of *the which* being used mostly in prepositional phrases in the early 16th century (see example [11] below). *Which*, on the other hand, was preferred as the subject of the relative clause (212).

When looking at the historical development of *the which*, it becomes clear that further research is, indeed, much needed. The debate begins with its very origin, as its initial rise and spread remain mostly unclear. On the one hand, the pronoun presents close resemblance to

[3] Relative determiners typically determine a noun phrase that links to a relative clause. Apart from *which*, in Present–day English this function is often performed by *what*, *whatever* or *whichever* (Payne & Huddleston 2002: 398).

French *liquels*, suggesting a correlation between the two. However, *the which* first appeared in northern texts and slowly wound its way south during the 14th century. This makes French origin unlikely, given that French influence was generally weaker in this part of the isle. Alternatively, its origins have been suggested to be native and the first element *the* descending from the demonstrative pronoun that, originally, played a significant role in asyndetic relative clauses[4] (Curme 1912: 195–200; 355–72; Reuter 1937; both cited in Raumolin–Brunberg 2000: 210; Fischer 1992: 303; see also Dekeyser 1984: 61). Similarly debated is its rapid disappearance. *The which* fell into disuse during the early 1600s, with no instances appearing in the Helsinki Corpus in the second half of the century (Rissanen 1999: 297).

Generally, *the which* seemed to be more frequent in non–restrictive continuative clauses[5] than in other clause types (see example [12]). This preference is most likely motivated by the need for a clear reference in the case of longer, sentential antecedents. *Which* or *that* may, in contrast, have caused ambiguity in this role. Therefore, it also makes sense that *the which* appeared far more frequently in prose than in verse. Particularly, it was mostly used in academic text types, as these genres are more likely to contain longer embedded sentences that convey strings of information to the reader (Fischer 1992: 303–4).

In example [11], *the which* is linked to the main clause by a preposition (underlined). Example [12], on the other hand, features the relative pronoun in a continuative relative clause:

[11] Let the water be boiled, <u>in</u> *the which* one may put when it bubbles up a little cinnamon, […]. (Jean Liébault, 17th century. *MS Hunter 303*. MCEModESP)

[12] And againe set it over the fire with á moderate heat and gently to boyle. Till it be in the forme of á plaister, *the which* you may well knowe by often dropping it on á colde place of wood. (Unknown author, early 17th century. *MS Wellcome 8086*. MCEModESP)

Romaine's (1982: 71) claim that the *wh*–pronouns were preferred in higher registers, furthermore, suggests that the relativiser may have been dominant in more formal contexts.

[4] An asyndeton refers to a "form of verbal compression which consists of the omission of connecting words […] between clauses" (*The Oxford Dictionary of Literary Terms* 2015). An asyndetic relative clause is, therefore, linked to the main clause without a conjunction.

[5] In Rissanen's words, "in the discussion of the spread of the *wh*–forms, it has proved useful to distinguish a special type of non–restrictive clause called 'continuative'. In this type the two clauses stand in coordinating rather than subordinating relationship" (1999: 293). Modelled on Latin, continuative relative clauses can introduce a new sentence or paragraph and were especially frequent in the 16th century (Nevalainen 2006: 86).

Keeping this in mind, it can also be speculated that *the which* may have been less present in oral language or, perhaps, even exclusive to writing. This is supported by the data from Johansson's (2012: 785) study on Early Modern English relativisers, which only featured a few sporadic occurrences of the pronoun in transcripts of trial proceedings (taken from Kytö & Culpeper's (2006) Corpus of English Dialogues 1560–1760 (CED)). More instances, instead, appeared in drama texts containing planned language.

2.1.5 Previous studies

As mentioned in section 2.1.4, studies focussing on the relativiser *the which* are few. General surveys and corpus–based studies of the Early Modern English relative system have built a diachronic overview of its development, studying its first appearances in Middle English texts and its quick disappearance during the 17th century (see, e.g., Curme 1912; Reuter 1937; Mustanoja 1960; Rydén 1966; Fischer 1992; Rissanen 1999; Raumolin–Brunberg 2000 and Nevalainen et al. 2011). Furthermore, conclusions have been reached on the text type *the which* mostly appeared in and its patterns of use. The various accounts, however, present an important issue: observations made by scholars and their following generalisations often seem to contradict each other. For instance, the actual frequency of *the which* is still largely debated. Mustanoja (1960: 199) claims that in 15th–century prose the combined pronoun was generally more popular than plain *which* (with exceptions that can be traced back to authors' individual style). This statement, however, is not supported by the data in Raumolin–Brunberg's (2000) or Nevalainen et al.'s (2011) studies. The latter even claim that *the which* was, most likely, always the minority variant in the wider language community (10).

Regarding the way in which the relativiser was used, Raumolin–Brunberg (2000) observed the trend that *the which* was more often used in prepositional phrases in the early 16th century, while *which* was preferred as the subject of the relative clause. This account, however, contradicts Reuter's conclusion that *the which* was, at all times, less frequently used after a preposition (1937: 176). Finally, Curme stressed that *the which* was the preferred choice in official, administrative documents (1912: 199). Fischer, however, convincingly argues that the main text genre it appeared in was didactic prose (1992: 303–4). In summary, only few conclusions on the use of *the which* are commonly accepted, and several speculations from earlier studies have been refuted over time.

In this project, I paid particular attention to the claims by Romaine (1982), Fischer (1992) and Raumolin–Brunberg (2000). Their speculations about *the which* and its uses should, of course, not be accepted without further questioning, given the contradicting accounts on the topic. Rather, their conclusions were used to form a theoretical basis for this study, as they provided important indicators of the relativiser's distribution (see also chapter 3).

Romaine's (1982) and Fischer's (1992) main claims were introduced in chapter 2.1.4. While the latter convincingly illustrated the development of *the which* and its contexts of use, Raumolin–Brunberg has conducted the so far only large–scale corpus–based study on the pronoun, that I am aware of. In her (2000) paper, she investigated and compared the distribution of *the which* and *which* in the Helsinki Corpus and the Corpus of Early English Correspondence (Nevalainen et al. 1998). The timespan she focused on was 1417 to 1500, coinciding thus with the Late Middle English period (in addition, she included the first Early Modern English subsection of the Helsinki Corpus in her study). Her main aim was to investigate whether *which* and *the which* acted as a free variants, or whether there were fine differences in their use. Her study did not produce any significant new conclusions, as her analysis was complicated by the extensive textual variation in her data (see section 3.1). Despite not spotting any clear patterns, her article served as an important model for this project.

Her paper was also used as a basis for Nevalainen et al.'s (2011) study on the linguistic progressiveness and conservativeness of individuals in Early Modern England. Among other features, they studied the distribution of *the which* across the Corpus of Early English Correspondence. Their main observation was that the wool merchant community of London seemed to favour *the which* over plain *which*. Authors from other regions and with a different social background, on the other hand, did not (14; 34). At first glance, this seems like a new conclusion that adds another context to the text types *the which* typically occurred in, besides academic prose. However, this finding may be related to the dataset that was used. Raumolin–Brunberg, using the same corpus and, consequently, obtaining the same result, pointed out that this speaker community was overrepresented in the data. The conclusion that *the which* was the dominant form among professionals should be handled with caution, as there is not sufficient data to extend this claim to any other groups of merchants or other non–gentry (2000: 218).

2.2 Early Modern English medical prose

The main genre that *the which* is believed to have appeared in is that of academic prose. This is not surprising, as *the which* might have been the preferred choice in continuative clauses where a clear link to the antecedent was needed to avoid ambiguity. This type of relative clause may be expected in texts of a higher register, rather than in writing representing spoken language (Fischer 1992: 303–4). In addition, as observed by Romaine, the *wh*–pronouns were generally more popular in formal registers and not as present in everyday colloquial language (1982: 71). These characteristics, which were at least partially confirmed in Raumolin–Brunberg's (2000) data, led me to focus on the related genre of medical writing in this project.

In the following subsection, I will focus on Early Modern English medical literature as a focus of historical and sociolinguistic research. Special attention will be paid to the genre's linguistic properties. In section 2.2.2, I will provide an outline of the main Early Modern English medical text types. Finally, in section 2.2.3 the focus shifts to society. Here, I will discuss the availability and accessibility of medical writing in Early Modern England.

2.2.1 Early Modern English medical writing as a research focus

Especially over the last two decades, medical writing has been the central focus of numerous studies. Pahta & Taavitsainen suggest that the reasons for this particular interest may lie in the texts' target audience, as well as the societal insights they provide. When looking at medical texts it does, in fact, become clear that the main aim was no longer to solely reach the elite. Instead, the readers formed a heterogeneous group composed of middle–class merchants, craftsmen, women, and the poor. This diversity, in turn, becomes especially interesting in the fields of historical and sociolinguistics, as the data reveals insights into diverse language use and communicative strategies (2011: 5). Consequently, research in Early Modern English medical writing approaches the genre from a variety of angles. Studies may be corpus–based or corpus–driven[6], make use of advanced statistical tests, and cover topics such as historical pragmatics, semantics, and, typically, language variation and change (see, e.g., Taavitsainen & Pahta 2011).

[6] Corpus–based research assumes the validity of linguistic forms from linguistic theory. Corpora is employed to investigate and further analyse them. In corpus–driven research, on the other hand, the linguistic structures themselves emerge from the corpus analysis and determine the further development of the study (Biber 2009: 159).

Another indicator for the increasing popularity of medical texts in linguistic research is the emergence of medical corpora. The probably leading dataset is Taavitsainen & Pahta's (1995–) Corpus of Early English Medical Writing (CEEM), which counts 2 million words in its Early Modern English subsection and offers rich metadata about its texts and authors (Pahta & Taavitsainen 2011: 16). Another noteworthy corpus is Calle Martín et al.'s (2012–) Málaga Corpus of Early English Scientific Prose (MCEESP), which was used in this study (see section 3.1.2).

Early Modern English medical writing presents a rich variety of text types, written for different purposes by authors from different social and educational backgrounds. The differences between professionals' and laypeople's language are often directly reflected in their writing. For example, McConchie & Curzan's (2011) study revealed that practitioners with a university education were generally more precise and innovative in their texts. Likewise, as mentioned above, the target audience also represented all social ranks (Pahta & Taavitsainen 2011: 4–5). Claiming that medical writing presented a high and formal register as a whole would, thus, be an overgeneralisation. This, however, is the reason why the socioeconomic background of the authors and readers forms an important variable in research. After all, as observed by Bell, "the audience a speaker/writer has in mind can significantly influence their speech and writing" (1984; cited in Barth & Schnell 2021: 16). Within the context of this study, this means that *the which* may only be present in texts written by authors with a higher professional background that wrote for educated readers.

One of the most important linguistic aspects in Early Modern English medical writing is the vernacularisation from Latin to English. In the previous centuries, the institutional language in any field related to the sciences or academia was either Latin or, to a lesser extent, French. Within this diglossia, these languages were used as prestige varieties, while English represented the informal, low variety (Nevalainen 2006: 29). Voigts points out that, in the preceding centuries, the prevailing of Latin affects research to the extent that "[working] with Middle English medical writing is, […], to work with Latin medical writing" (1984: 316). The Early Modern period, however, saw a gradual shift, and by 1700 English had clearly replaced Latin as the dominant language in scientific literature (Pahta & Taavitsainen 2011: 4; see also Crossgrove 2000).

This being said, medical publishing was highly conservative at the beginning of the period. Printers preferred to distribute established Latin titles that had been well–received in

the past. When English began to gradually replace Latin, as much as one third of the works published in the 16th century were translations (Taavitsainen et al. 2011: 12–13).

Despite the initial conservatism, the change from the previous institutional language to the vernacular was a turning point in the history of English. Scientific and academic literature in English meant increased accessibility for the wider population, that was amplified by the thriving of the printing and publishing industry. Nevertheless, as pointed out by Slack (1979: 258–61; cited in Jones 2011: 38), this does not imply that English medical texts should be automatically regarded as "popular medicine". In fact, several factors, such as their pricing and their prefatory material, suggest that in the case of elaborate medical books the target audience rather included the gentry, wealthy merchants, and the middle class. Instead, recipe collections, often written by laypeople, or almanacs were reserved for the lower ranks of society (see section 2.2.2).

In the context of Early Modern English and the ongoing standardisation processes, scientific writing presents a special case. Compared to other genres, the general spread of a standard spelling system was slower in this text type. This can, at least with works composed by qualified professionals, be motivated by their intellectual prestige and higher register that demanded more conservative language use (Taavitsainen 2000: 147; see also Schnell 2009: 4). This makes scientific texts comparable to other formal genres, such as legal writing. The latter, typically presenting formulaic language, is also known for adhering to standardisation processes at an own pace and for undergoing similar developments, including the switch from Latin to the English vernacular. On the other hand, because of their level of formality, these genres may have been especially committed to a regularised spelling and the removal of regional variation (Kytö et al. 2007; cited in Meurman–Solin 2012: 672). Consequently, the diachronic development of relative clauses in medical writing forms a valid research focus for this project. Particularly interesting are aspects such as the disappearance of *the which*, or innovations like the animacy parameter and when they took place. Based on the characteristics mentioned above, there are grounds to believe that these developments may have been completed at a different pace than in other genres.

2.2.2 Medical text types in Early Modern England

In general, the boundaries between strictly medical texts and those belonging to the natural sciences can be fuzzy in the Early Modern period. A clear classification of a given text

is, therefore, not always possible (Early Modern English Medical Texts (EMEMT)). Traditionally, scientific texts have been divided into the classes of 'theoretical texts', 'surgical texts' and 'remedies' (Voigts 1982; see also Taavitsainen & Tyrkkö 2010; cited in Calle–Martín 2021: 116). The three categories have different degrees of linguistic complexity and can be further divided based on their level of prestige and formality. Theoretical texts are considered the most academic, as they were mainly intended as accounts and handbooks of qualified practitioners for fellow scientists and doctors. Remedies, on the other hand, were mostly directed at a lay audience and would adopt a more colloquial tone, avoiding excessive medical jargon. In cases like these, the medical formation of the authors themselves may also be questioned, as it was not uncommon for laypeople to write remedy and recipe books. Surgical treatises would fall in–between the two classes, as surgeons frequently covered a broad variety of topics in their works. Surgical texts could, for example, also contain recipes or cures for external discomforts (Voigts 1982; Pahta & Taavitsainen 2004: 7; cited in Calle–Martín 2021: 116–17; Taavitsainen et al. 2011: 24; see also Romero–Barranco 2020).

Despite the evident differences between these three groups, medical writing in general still represents a highly specific context. Compared to other genres, the typical language in scientific texts is more formal and serves an informative communicative function, particularly when composed by authors of a professional background. Especially after the institutionalisation of medicine and the rise of scientific professions in society, less colloquial and more academic writing conventions are to be expected (see Hiltunen & Tyrkkö 2011: 45).

In his (2020) study on linguistic complexity across medical text types, Romero–Barranco analysed the language used in a surgical treatise and a collection of medical recipes (falling into the category of 'remedies'). His conclusions confirmed the generally described characteristics of the two text types, as the surgical treatise indeed presented more of the features that typically indicate linguistic complexity. He also observed that pied piping relative clauses were frequent in this work, serving as a further indicator of prestige[7]. However, the recipe collection also contained a significant number of features associated with lexical specificity, and the use of participial clauses. This shows that even in less formal texts that were directed at a lay audience, medical writing had certain conventions associated with a higher register, that authors would generally follow (67–8).

[7] According to the *Oxford Dictionary of English Grammar*, pied piping refers to the "placing of a head and its associated complement (typically a preposition and its complement) at the beginning of a clause". Traditionally, pied piping is associated with more formal registers (see, e.g., Hoffmann 2011: 1; 3).

It was also common for medical books (and Early Modern English books in general) to contain introductory dedications, verses praising the author, recommendations of colleagues, or words directed at the reader himself. These paratexts, designed to influence and control the audience's reading, were indispensable until the mid–17th century. Afterwards, they began to be considered old–fashioned or inappropriate by some critics (Jones 2011: 42). In general, scientific writing in Early Modern England employed a number of particular communicative strategies. The focus of elaborate introductory paragraphs and personalised comments was, in most cases, the relationship between the author and the reader. People were becoming increasingly interested in medicine, and scientists and doctors were highly respected in society (Jones 2011: 38–9). Apart from the audience involvement, changes in the politeness scale were also apparent, and authors made increasing efforts at presenting themselves as highly educated and eloquent professionals (Pahta & Taavitsainen 2011: 5).

2.2.3 Accessibility of medical texts in Early Modern England

The way in which science is communicated is directly linked to a society's ways of scientific thinking (see Taavitsainen & Pahta 1995; cited in Pahta & Taavitsainen 2011: 3). The Early Modern period saw numerous important changes, both from a societal and a linguistic perspective (see section 2.1). It is not surprising that medical prose at the beginning of the 16th century still largely favoured traditional patterns, closely following the models and conventions suggested by previous texts. In the following decades, however, medieval science that based itself on the theories of ancient writers, such as Galen and Hippocrates, was increasingly being questioned. By 1700, scientists and doctors clearly preferred empirical methods and founded their knowledge on experiments and observation (see Crombie 1994 and Taavitsainen & Pahta 1995; Pahta & Taavitsainen 2011: 3–5). Simultaneously, science and medicine's modes of communication also underwent developments. By the middle of the 17th century, a large variety of publications was available, including textbooks, treatises, health guides, and remedy collections (Taavitsainen 2010: 47). In addition, printing had made texts available to a wider audience of readers. Particularly works providing basic medical knowledge for laypeople, such as almanacs, were also sold at relatively cheap prices (Taavitsainen et al. 2011: 9).

However, an important aspect that needs to be kept in mind is the general literacy of the Early Modern English population. The degree to which individuals were able to read and, likewise, the audience printing houses had in mind, has been keenly debated by historians. While writing leaves at least a few clues on people's literacy, it has proved virtually impossible

to statistically measure their reading abilities. In the case of writing, a method that has been adopted is that of studying signatures, since in Early Modern England the signing of documents had become a regular practice. The subscriptions to the Protestation Oath of 1642 showed that 70% of the adult English men and 90% of the English women were illiterate (understanding signing as the full spelling of their names, as opposed to leaving their marks on documents) (Cressy 1980: 176–7; cited in Jones 2011: 30). Based on Cressy's methods, it has also been measured that male illiteracy declined from around 90% in 1500 to 55% in 1714. On the other hand, 75% of all women in 1714 were still illiterate (1980: 176–7; cited in Jones 2011: 31). Of course, this measure of subscriptional literacy only gives estimates about individuals' writing abilities. Its reliability is also compromised by the overrepresentation of the clergy and higher ranks of society. The female percentage of the population and the working class, on the other hand, are severely underrepresented. Furthermore, it is not a suitable method to also investigate people's reading capacities. Reading was normally taught in schools before writing, and it can be assumed that many individuals, who had acquired a basic level of education, may have been competent readers. However, they may have never adopted the practice of signing with their name and, instead, only made their mark on documents. Also, instances are known where persons chose to sign with their mark instead of their name, despite being able to do so. As a consequence, any estimated numbers regarding the literacy in Early Modern England are, indeed, only indicators and have no certainty (Jones 2011: 31).

A direct correlation can, naturally, also be observed between people's degree of literacy and their position in society. Based on their ability to sign documents, all clergymen, noblemen and members of the professional classes in 17th–century England could read and write. On the other hand, 85% of labourers and high numbers of servants are estimated to be illiterate (Cressy 2003: 315; cited in Jones 2011: 30–1). Furthermore, there was also significant geographical variation and, especially, the contrasts between country and town were striking. In London, for instance, only 22% of adult men were measured as illiterate based on their subscriptional competence in the mid–1600s (Jones 2011: 31).

It seems, thus, that medical texts and literature in general were only available to a limited portion of the population. While there must have been a noticeable increase in people's ability to read and write between 1500 and 1700, the access to medical writing might largely have been reserved to the higher, educated ranks of society. And yet, treatises, recipe collections, almanacs and preventative health guides were aimed at a wider audience, including laypeople and the lower ranks of society (see Taavitsainen et al. 2011: 23–4). This becomes particularly evident

in texts written in dialogue form, in verse, or as sayings and proverbs that served mnemonic functions. Several texts, such as proclamations, plague orders or bills of mortality, were often clearly meant to be read out loud in a public setting, for example in the local market square or on the main streets of towns. In cases like these, the illiterate percentage of the population was still able to access information and news (Jones 2011: 33–4).

2.3 Language change and the individual

After outlining the linguistic situation in Early Modern England and introducing the genre of medical writing, the focus in this section shifts to the people who lived, spoke, and wrote between 1500 and 1700. One of my main aims in this project was to investigate the distribution of *which* and *the which* with regards to the individuals who used them (or who, perhaps, chose not to use *the which* at all; see chapter 1). In the next subsection, I will introduce the concept of linguistic lifespan change, followed by the special case of Early Modern English and its speakers in section 2.3.2.

2.3.1 Linguistic lifespan change research

Language does not exist without language users, and changes cannot take place and spread without interactions between them (see, e.g., Hopper & Traugott 2003: 40 or Raumolin–Brunberg & Nurmi 2011: 252). A crucial factor that, all too often, has been overlooked in studies concerning language change is the role of the speakers. The question of how ongoing shifts and innovations influence people's choice of language, and how they are reflected in individuals' speech and writing, is a central part of this study.

It is widely known that language users continue to expand their vocabulary well into adulthood (see, e.g., Brysbaert et al. 2016), but the issue of grammatical changes is debated. A reason for this may be the traditional tendency in grammaticalisation studies to focus on generational change, rather than change in an individual speaker's language (Anthonissen & Petré 2019: 1).

The generative approach bases itself on the hypothesis that the youngest generation, i.e., children acquiring a given language as their mother tongue, reanalyses ambiguous linguistic input and is the main driving force behind linguistic change. The assumption is that each human possesses innate universal capacities of language use and language processing, that enable them

to form and internalise a set of rules (grammar). Their output, in turn, will influence the following generation, that will reanalyse the perceived patterns. The next generation's grammar may, therefore, present differences from its predecessor (see, e.g., Andersen 1973: 778; Anttila 1989 [1972]: 197; Lightfoot 1979; Hopper & Traugott 2003: 40–1; Anthonissen & Petré 2019: 2 or Fonteyn & Petré 2022: 79–80).

Despite generativist views counting among the most influential theories of language change, they may not account for some important aspects. For example, in the context of a heterogenous, multilingual speaker community, the role of social factors should not be underestimated (see, e.g., Labov 2001). Simultaneously, generativist models seem to imply that the grammar and the output of an individual remain mostly static over the course of their life. While it is certainly true that the main linguistic learning processes take place during the early stages of childhood, it is unlikely that a person's language remains unvaried for the rest of their life.

A different approach is taken in usage–based models. Here, the focus lies on both the language of the individual and the speaker community, and it is assumed that language is a dynamic and "adaptive" system (Diessel 2019). A person's grammar is, thus, directly influenced by other speakers and social interactions with them (Bybee 2006; Barlow 2013: 444; all cited in Fonteyn & Petré 2022: 79).

As pointed out by Fonteyn & Petré, however, "the two types of models are not always at odds" (2022: 80). For example, there is consensus between the two approaches that lexical or so–called "surface structure" changes may be well represented by the adult population. An example would be the replacement of the third person singular suffix –(*e*)*th* by –(*e*)*s* in Early Modern English. This change was the focus of Raumolin–Brunberg's (2005) study, who analysed letters of adult speakers (Fonteyn & Petré 2022: 80). Her research aims present similarities to this project, and there are grounds to believe that changes may also be observed in the gradual disuse of *the which* (see section 2.3.2).

Apart from innate linguistic abilities, (second) language acquisition has especially in the last decades been attributed to brain plasticity. Once the critical period in a person's (early) life reaches its end and the plasticity declines, so does language learning ability (see, e.g., Birdsong 2005; Rowland 2014; Birdsong & Vanhove 2016)[8]. Based on this theory, the situation may be

[8] In Krulatz et al.'s words, the critical period can be defined as a "hypothesised period during childhood when [a person has] to acquire a language in order to acquire native–like competence" (2018: 41).

similar with the grammatical aspects an individual internalises during childhood. Consequently, a person's abilities to change and/or replace elements of their native language should decline with age (Sankoff 2019: 197, *footnote 1*).

Sankoff (2018; 2019) defined three trajectory types that language change may take over an individual's life. In the first scenario, which is the most common, speakers may prefer to stick to the patterns they acquired early in their lives. Thus, they do not actively participate in ongoing language changes. This trajectory would seem to confirm the generativist models. The second possibility, however, is that older speakers may adopt new features that are used by younger generations, accelerating the spread of the change, and replacing old speech patterns. Another possibility is the third scenario, which foresees a conservative behaviour of the speakers. In rarer situations, especially aging speakers may choose to readopt older patterns, showing retrograde lifespan change and moving against the speech community. All trajectories are influenced and, to a degree, delimited by brain plasticity. However, social and cultural factors also play an important role.

It seems, thus, that the assumption of language change being carried consistently from one generation to the next would be an oversimplification. Instead, an interesting research focus would be the factors contributing to one trajectory taking place rather than another.

Generally, the behaviour of the individual (adult) speaker has not been the focus of many projects, at least in the context of grammaticalisation and language change (Raumolin–Brunberg & Nurmi 2011: 251). However, the topic has sparked particular attention in recent years, resulting in an increasing number of studies (e.g., Sankoff 2018; 2019; Anthonissen 2019 or Fonteyn & Petré 2022) and corpora designed specifically for this purpose. The leading corpus for quantitative studies on linguistic lifespan change, EMMA (Petré et al. 2018), will be introduced in section 3.1.4.

In this study, a special aim was to investigate the language of individual Early Modern English authors. With *the which* gradually disappearing during the 17th century, and within the context of standardisation processes, a valid research question was how the ongoing changes were reflected in the writing of persons who experienced them first–hand.

2.3.2 Linguistic lifespan change in Early Modern England

In the first sections of this chapter, I outlined the linguistic situation in the Early Modern English period and the most important changes that took place in it. The main focus in section

2.2 was the rise and spread of medical literature, which was facilitated through the thriving of printing houses. A crucial question that has only briefly been mentioned concerns the people who lived in Early Modern England: how did the ongoing changes affect their language? The years from 1500 to 1700 are characterised by processes that resulted in a written standard language. There is no doubt that the way people wrote and spoke must have been directly affected by them, at least to some degree.

While the Early Modern English period is well–researched and diachronic investigations are numerous, corpus–based studies with the individual in mind are few. Even rarer is research on the way persons' language may have changed over the course of their life. A highly relevant study was, however, carried out by Raumolin–Brunberg. In her (2009) article she studied the lifespan changes in the language of three early modern gentlemen. Analysing multiple letters contained in the Corpus of Early English Correspondence (CEEC), which were composed over several decades, she argued that the authors' grammar was indeed not fixed. Instead, the ongoing linguistic changes were directly observable in their writing.

A similar study was carried out by Raumolin–Brunberg & Nurmi in 2011. Their focus lay on the grammaticalisation of the auxiliaries *do*, *will* and *would* in Early Modern English. To track their development, personal letters (also taken from CEEC and its Supplement (CEECSU)) of three contemporary adult authors were analysed. Their most interesting finding, perhaps, was the difference between the investigated features. While in some cases the subjects adopted the new forms and meanings, in others they preferred to stick to the patterns acquired during their youth for their entire writing. It seems, thus, that previous conclusions on linguistic lifespan change were confirmed in their data. Their language did continue to evolve during adulthood, providing evidence against the claim that a person's grammar is fixed after the early years of childhood (see, e.g., Petré et al. 2019 or Fonteyn & Petré 2022). Furthermore, direct parallels were observed with the trajectory types defined by Sankoff (2018; 2019).

In this project, a special focus lies on the way *which* and *the which* are distributed in individual authors' writing. Especially the quick disappearance of *the which* is an interesting case, and part of the investigation concerns whether it is directly reflected in people's language. On the one hand, one may expect especially individuals with a higher social rank and a higher educational background to comply early with standardisation processes. After all, a principal aim of the processes was to limit variation and to form an official written standard, suitable for formal contexts. Individuals using the emerging standard, in opposition to dialectal varieties, were considered sophisticated and polite (see Stein 1993: 8; see also section 2.1). On the other

hand, previous linguistic lifespan studies (e.g., Raumolin–Brunberg & Nurmi 2011) found that some individuals prefer to stick to patterns acquired early in their life (see above).

The standardisation processes that took place between 1500 and 1700 can be considered language change from above. This type of linguistic change is of planned and artificial nature, and language users are generally aware of the ongoing processes. This also means that, ultimately, changes need to be accepted and adopted by a speaker community to fully establish themselves (Coates 1993: 169; see also Labov 1972). Even if people in Early Modern England were aware of the ongoing processes and changes, the degree to which their language was affected may have varied. Individual variation, depending on personal preferences and style, as well as social factors may have determined whether new features and uses were present in a person's language, or not. Time also plays an important role, as linguistic change is gradual. Consequently, certain features might rather appear in the language of individuals who lived towards the end of the period, when standardisation processes were slowly reaching completion.

Closely related to this are the notions of linguistic knowledge and language use. An individual's linguistic knowledge can be defined as the abstract system that comprises their complete linguistic repertory. Language use, on the other hand, refers to the elements and patterns which they actually include in their speech and writing (Barth & Schnell 2021: 7). For example, *the which* disappears from most Early Modern English corpora during the 17th century. However, authors were most likely still aware of its existence and contexts of use. It may have been associated with certain connotations, such as perhaps being old–fashioned or particularly sophisticated, given its association with high, academic registers (Fischer 1992: 303–4). Of course, it is not possible to make any conclusions on subjective matters like these, unless they are explicitly expressed in contemporary texts. However, the possibility that *the which* was perhaps still (at least to some degree) present in individuals' lives cannot be excluded either, as its disappearance was a gradual process and did not take place overnight. One of my main aims in the empirical part of this study was to look for indicators of this.

2.4 Summary

The aim of this chapter was to outline the setting of my project. In sections 2.1 to 2.1.5, I introduced the social and linguistic situation in Early Modern England. The period between 1500 and 1700 was characterised by important changes resulting in a standard variety. The main

developments concerning the relative pronoun system were the animacy distinction and the gradual disappearance of combined *the which*, which form the focus of this study.

After introducing relative clauses and their main pronouns, I moved on to the genre of medical writing in section 2.2. Scientific and medical texts bloomed in the 17th century and present a special case in Early Modern English literature. With a high register and prestige, medical writing can be expected to comply early with standardisation processes. On the other hand, it has been noted to follow an own pace with certain developments and to be rather conservative (Taavitsainen 2000: 147). This may, for instance, indicate a later use of *the which* than in other genres.

In the last section, 2.3, I talked about the field of linguistic lifespan change, which plays a central role in the second half of this project. In recent years, research (see, e.g., Bybee 2006; Raumolin–Brunberg 2005; 2009; Barlow 2013; Sankoff 2018; 2019 or Fonteyn & Petré 2022) has shown that a language user's grammar is by no means fixed. Instead, it is directly influenced by interactions and other social factors, and may likely evolve over time. A central focus of this study, therefore, is to investigate whether the changes in the relative pronoun system are visible in Early Modern authors' writing.

3. Methodology

In this chapter, I will discuss the methodology employed in the study in detail. Firstly, in sections 3.1 to 3.1.7, I will present the corpora and the tools used for the data collection. Particular attention is paid to their suitability and relevance for the project. In section 3.2, I will focus on the analysis of the data and the statistical tests that were conducted.

3.1 Corpora used in the study

A first step in designing and planning the study was to gain insight into previous research, that is into theoretical literature and corpus–based studies. With relativisation being one of the most studied topics in the history of English (Nevalainen & Raumolin–Brunberg 2002: 109), it was not a difficult task to gather enough material to form a theoretical basis for the project. However, as seen in section 2.1.5, contradictory conclusions posed a number of challenges when it came to formulating the research questions and hypotheses for this study.

Once the main aims of the project had been defined, the next step was to identify the dataset that should be used to conduct its empirical part. In other words, it was necessary to identify where to look for instances of *which* and *the which* to obtain interesting and valid results. Once again, I consulted previous literature for clues on the timeframe and text type(s) where the pronouns could be mostly expected. Generally, scholars agree that *the which* was common in the 16th century and its use started to decline in the following decades. Eventually, it fell into disuse in Modern English (Fischer 1992: 303–4; Rissanen 1999: 296–7; see also section 2.1.4). Keeping this development in mind, I decided to work with a text corpus covering the time span of 1500–1700, coinciding thus with the Early Modern English period. However, a general corpus containing virtually any genre and situational features might not have provided any significant data to analyse. As mentioned in the past chapter, previous studies have shown that *the which* was used predominantly in didactic prose and the *wh*–pronouns were, in general, associated with a more formal register (Romaine 1982: 71; Fischer 1992: 303–4; see sections 2.1.3.2 and 2.1.4). For this reason, a special corpus containing texts of a scientific genre seemed especially suitable for this study, to investigate patterns and uses of the relativisers in a highly specific context (see Cheng 2012: 32–4 or Barth & Schnell 2021: 34).

In the next sections, I will introduce the datasets that were eventually selected and the methods that were employed in the data collection. Before, however, two issues pointed out by

Raumolin–Brunberg (2000) in her study on whether *which* and *the which* acted as free variants need to be addressed. Firstly, after carefully comparing past studies, she observed that in many of them the resulting conclusions may have strongly depended on the corpus that had been used. Consequently, they often contradicted each other, and overarching generalisations were difficult to establish (215). For this reason, I decided to use more than one corpus in this study, as an attempt to avoid such limitations. The second aspect she noticed was that the data from the Helsinki Corpus, which she worked with (alongside the Corpus of Early English Correspondence), presented extensive textual variation. The fact that unequal amounts of writing from multiple different genres were included made it difficult to spot any patterns and to reach universal conclusions in her investigation (2000: 215). This limitation is, of course, not true for all studies that have been published on relative pronouns; Sato (2016; 2019), for instance, limited the dataset in her studies to Shakespeare's plays and investigated his use of the *wh*–pronouns with special attention to the animacy parameter and speakers' emotional motivations. This special focus proved to be fruitful, as she observed a correlation between Shakespeare's use of the relative pronouns and the *you/thou* paradigm. For this reason, limiting the data of this study to one genre seemed promising.

3.1.1 Selecting the datasets for the study

With a vast selection of articles and corpus–based studies on relativisers (see, e.g., Rydén 1966 or Dekeyser 1984; see also section 2.1.5), there was no shortage of models to follow for the empirical part of the project. Some studies (e.g., Raumolin–Brunberg 2000, Nevalainen & Raumolin–Brunberg 2002 or Johansson 2012) also had a special focus on extralinguistic factors, and their influence on authors' use of relative pronouns was investigated. In cases like these, therefore, a sociolinguistic perspective was also taken. While this is also the case in the second part of this study (see section 3.1.3), I took a more traditional quantitative approach in the first half.

As mentioned in section 3.1, I decided to work with corpora containing texts of a didactic type. More specifically, the focus lay on scientific writing. A first choice fell on the Early Modern English component of Taavitsainen & Pahta's (1995–) Corpus of Early English Medical Writing (CEEM); however, the idea was soon discarded due to technical issues which made it impossible to work with the software. The 2–million–word subcorpus offers rich material of the medical genre, includes extensive metadata, and aims at fulfilling a

representative function of the medical literature produced between 1500 and 1700 (Taavitsainen et al. 2011: 16). Furthermore, the corpus is accompanied by a book with a considerable number of studies conducted with its aid (Early Modern English Medical Texts (EMEMT)). For this reason, it had presented a natural first option. On the other hand, the research conducted with it has, so far, rarely centred around syntactic issues and the corpus has been preferred for lexical studies (see e.g., Mäkinen 2011 or Ratia & Suhr 2011). Paired with the previously mentioned technical issues, I concluded that a different dataset might be more suitable for this study's particular concerns.

3.1.2 The Málaga Corpus of Early English Scientific Prose (MCEESP)

Eventually, I considered the Early Modern English component (MCEModESP) of Calle Martín et al.'s (2012–) Málaga Corpus of Early English Scientific Prose (MCEESP) to be particularly fitting for the project. Comprising a little over 1 million words, the subsection containing medical texts from 1500 to 1700 was smaller than EMEMT, but comparable in terms of the timespan covered and the text genres. The source manuscripts contained in the corpus stem from the Hunter Collection at the University of Glasgow Library, the Wellcome collection at the Wellcome Library in London, and the Rylands collection at the University of Manchester Library, and present rich insights into the language used in, mainly, theoretical treatises, surgical treatises, and remedies. A further advantage was the corpus' accessibility, as it is openly available and can be downloaded directly from its official website. In addition, the corpus comes in the formats of its plain, normalised, and part–of–speech–tagged text. Its plain version contains the semi–diplomatic transcriptions of the original texts, preserving their original spelling and word division. In the normalised format, the variant forms were standardised to Present–day English using the VARD software. The part–of–speech–tagged version, which was automatically annotated with the Constituent Likelihood Word–tagging System (CLAWS), was used for the data collection, as the tagging made a faster distinction between relative and interrogative pronouns possible (The Málaga Corpus of Early Modern English Scientific Prose 2007–2016; Calle–Martín 2021: 120). An exception here were the text files *MS Hunter 487*, *MS Wellcome 774*, *MS Wellcome 7818* and *MS Wellcome 8086*, as they were not available in the tagged version of the corpus and had to be included in their plain text format.

Since the downloaded files do not come with any integrated tools, I used the corpus analysis software AntConc (Anthony 2023) in its 4.2.2 version to collect the data. In its current release, the program includes effective tools to analyse collocations, keywords in context,

clusters, and lexical bundles. Word and keyword frequency generators, a word distribution plot and a wordcloud tool are also provided. The toolkit is, further, characterised by its user–friendly layout and can be freely downloaded from its official homepage (Anthony 2005: 729; 735; AntConc Homepage).

Frequency lies at the base of all aspects corpus linguistics are interested in and is, indeed, its most basic measurement. How frequently certain patterns are chosen over others, and in which contexts, reveals crucial information about the distribution and use of a certain feature (see McEnery & Hardie 2011: 28 and Barth & Schnell 2021: 44; 68–9; 71). Therefore, my first step consisted in generating wordlists by entering the relative pronouns *which*, *the which*, *who*, *whose* and oblique *whom* (paying special attention to spelling variants such as *whiche*) into AntConc's Word tool and noting their total occurrences. The texts were divided based on their (approximate) year of production and two separate groups were formed, namely texts from the 16th and texts from the 17th century. This division was made to make the diachronic development of the relativisers easier to observe. However, it needs to be noted that the two groups were not equal in size. The second group, containing all texts from the 17th century, counted a total of 1,128,878 tokens (including the tagging), while the texts from before 1600 only amounted to 124,698 tokens. The fact that the texts from after 1600 counted a million tokens more had to be taken into consideration during the later analysis, to make the resulting numbers comparable.

The next step was to enter the relative pronouns into AntConc's Keyword in context (KWIC) tool to investigate the clauses they occurred in. Here, special attention was paid to their individual antecedents. In the following two examples, the referents are underlined, with *which* in [13] referring to the noun phrase *the heart*, and *the which* in [14] to a sentential antecedent:

[13] To comfort <u>the heart</u> *which* is weak. (Bartholomew Blogate, c. 1650. *MS Hunter 64*. MCEModESP)

[14] […], <u>some Physicians give counsel to use some of the anacardic confection twice or thrice in the week</u>: *the which* I do much approve providing that it be done with the advice of the learned Physitian there present, […]. (Jean Liébault, 17th century. *MS Hunter 303*. MCEModESP)

Based on the available literature on the topic and previous research, it was expected that the relativisers would follow certain patterns in relation to their referents. For instance, *the which*

was expected to refer to sentential antecedents more often than *which* (see Fischer 1992: 303–4). At the same time, the mid–15th century data from Raumolin–Brunberg's (2000: 214) study suggested that *which* might have been preferred for singular antecedents. Whether this pattern was still common in Early Modern English was a matter worth investigating. In general, the colligational patterns[9] of the relativisers were expected to reveal interesting insights into their use and distribution. For this reason, their antecedents had to be collected and also classified. The categories I established for the antecedents depended on whether the referent was a noun phrase or a clause. In the former case, they also depended on the type of head noun in the antecedent noun phrase. The ultimate categories consisted of common nouns (excluding nouns that indicated a person; a distinction between singular and plural nouns was made in both cases), persons, deity, pronouns (a distinction was made between personal and nonpersonal referents), entire clauses and cases in which the relativiser acted as a determiner. In addition, a category labelled "others" contained any remaining instances in which the pronoun did not relate to a clear antecedent (for example, when the relativiser appeared in a free relative clause). Any unclear or unreadable cases which could not be clearly categorised were included in a separate class. To facilitate this operation, the default context size setting in AntConc's Keyword in context (KWIC) tool was increased from 10 to 25 visible tokens preceding and following the relative pronoun.

I quickly noticed that in the case of *which* and *the which*, the texts from the 17th century yielded too many hits to consider in their entirety. After an initial attempt to include them all in the study, it became clear that classifying and analysing each of the, respectively, 1,418 and 188 instances would have resulted in a highly time–consuming undertaking far beyond the scope of this project. Eventually, I decided to limit the number of hits of *which* and *the which* to a randomly generated sample of 100. To make the raw frequencies directly comparable and observe the diachronic development between the two centuries, I calculated the resulting percentages of each relativiser (see chapter 4 for the results).

Generally, at times the classification of the antecedents turned out to be more challenging than initially anticipated. This was mainly due to the clauses oftentimes being rather ambiguous and the antecedent being difficult to identify clearly. In certain cases, the tagging of the corpus and the wider context of the relative clause provided useful hints.

[9] Colligation, in contrast to collocation, does not refer to the co–occurrence of certain lexemes but the grammatical relationship between the parts of speech. For example, an article is expected show different colligational patterns with nouns, adjectives, and adverbs (see, e.g., Barth & Schnell 2021: 14–15 or Cheng 2014: 7)

Another drawback I soon noticed was that the metadata provided by the corpus was rather limited. While in the case of some files (e.g., *MS Wellcome 373* or *MS Wellcome 3769*) the author and the year of the original manuscript were known and noted, in others the author was unknown and/or the date of composition could only be estimated (in which case rather broad timespans such as, for example, "between 1499 and 1599" were indicated). As calculated by Fissell (2007; cited in Taavitsainen et al. 2011: 15), as many as 17% of vernacular medical books published in the second half of the 17th century were anonymous. It is, therefore, not surprising that the background information from, especially, earlier manuscripts was rather incomplete. The information that was contained in the files or available on the website of the respective library (see the official websites of the University of Glasgow Library, The Wellcome Library and The University of Manchester Library) was highly valued and taken into consideration during the analysis. However, given the wider scope of the study, it soon became clear that it was not sufficient, and more detailed information on the situational features of the texts (including insights into the socio–economic and educational background of their authors) was needed to answer the proposed research questions. After all, metadata on the extralinguistic features of the texts in a corpus is a central prerequisite for its basic representativeness and, further, determines the type of research that can be conducted with it (Barth & Schnell 2021: 33; see also McEnery & Hardie 2011: 29–30). Given that, in this particular study, I also wanted to focus on more sociolinguistic topics, it became clear that an additional corpus was needed for this purpose.

3.1.3 Sociolinguistic focus of the study

The lack of background information on the authors of the manuscripts in MCEModESP, as well as its rather limited size, made it clear that an additional dataset was needed. Furthermore, an improvement over Raumolin–Brunberg's (2000: 215) study, which was suggested by her to reduce textual variation, was achieved by using a corpus containing an exclusive text type. However, the valid point she made that past conclusions were limited to their own dataset was still an issue. Adding a second corpus to the study and potentially observing patterns valid for both did, therefore, become one of the project's main aims. Also, the following research question was proposed in chapter 1:

> *How are the changes in the Early Modern English relative pronoun system visible in individual authors' writing, and what role does their socio–economic background play?*

This question bases itself on the well–established fact that language users' choice of certain linguistic features over others, as well as the adoption of new forms and/or meanings, is often constrained by social factors. Furthermore, variation is expected between different contexts and registers (Raumolin–Brunberg & Nurmi 2011: 252–3). With the Early Modern English period being highly interesting in terms of linguistic change and standardisation processes, the language of individual authors with regards to their social background presents a valid research focus (see section 2.3.2). Working with MCEModESP, it had not been possible to pay any special attention to the writers of the original manuscripts, given the general lack of metadata provided by the corpus. A more sociolinguistic approach was, however, taken in the second half of the study.

3.1.4 The Early Modern Multiloquent Authors (EMMA) Corpus

Eventually, I decided to add data from the Early Modern Multiloquent Authors (EMMA) Corpus (Petré et al. 2018) to my project. Designed specifically for the quantitative study of linguistic change over the lifespan of individual authors (Petré et al. 2019: 85), this corpus allowed the study to take a decidedly distinct turn after analysing the data from MCEModESP.

The compilation and the final version of the corpus are a result of the ERC–funded research project Mind–Bending Grammars and offer unprecedented new possibilities in corpus studies (Petré et al. 2019: 85). Most corpora that are available, as is also the case with MCEModESP, are composed of a variety of texts by different authors within a set time window. Unless the focus lies on specific authors and multiple instances of their writing are included, a diachronic investigation of individual language use is rarely possible.

Here, EMMA presents a new concept. Instead of covering the 17th century by gathering portions of available written material, the focus lies on the authors of the texts themselves and their entire writing careers. While the corpus is by no means the first project to pay special attention to the individual, it still undoubtedly provides unprecedented resources and insights. Comparable corpora exist for other disciplines, such as language acquisition (e.g., the CHILDES Corpus), and the field of historical sociolinguistics also presents a number of carefully curated collections with the individual in mind (for instance the Parsed Corpus of Early English Correspondence (Nurmi et al. 2006) and its 18th century continuation, the Corpus of Early English Correspondence Extension (Nevalainen et al. 2000–…)) (all cited in Petré et al. 2019: 85). EMMA successfully counterbalances their main drawbacks, which in the first

case would be the lack of extensive diachronic data of adult language development and, in the latter, their often rather limited size. In fact, the corpus contains texts from 50 carefully selected authors, documenting their writing careers over several decades and amounting to a total wordcount of over 90 million (Petré et al. 2019: 85–6; 115).

The authors had to fulfil certain criteria to be included in the corpus. The first concerned the total amount of written material they produced over their careers. The minimum number of words was set at 500,000, which was reached by most. The second criterion focused on the distribution of the texts across their, ideally, long careers. To observe relevant diachronic change, it was necessary for the texts to be spread as evenly as possible over a timespan of several years. The last desirable characteristic concerned their link to the London society. Despite the metropole presenting a significant variety in terms of both its population and their language use, it is fair to assume that a certain sense of identity was formed among, especially, the elite members of society (see Archer 2000 and Nevalainen 2015; cited in Petré et al. 2019: 87). Ideally, all three criteria would have had to apply to the selected authors for them to be perfect candidates. In practice, this was not always the case and, instead, an optimal balance was the primary goal (Petré et al. 2019: 86–8). The fact that the authors in the corpus did, therefore, for the most part belong to an exclusive portion of the London population and had a certain (high) status fit the scope of this study particularly well.

Specialised corpora focussing on specific speaker communities are, of course, widely available, and often rich in metadata. However, it is more difficult to find historical corpora covering older forms of English with detailed information about their texts and authors, simply because it may not always be available (Nevalainen & Raumolin–Brunberg 1989: 68; Barth & Schnell 2021: 265).

A vast amount of metadata is, however, available in the case of this corpus. Apart from the individual authors' names, years of birth and death, professions and connections between each other, an Excel sheet with extensive information about the texts and their properties (including their title, year of production and publication, publishing company, length in terms of tokens (estimated counts of Latin or French words are subtracted from the total number) and genre) is also provided. Additionally, the authors are further divided into five categories labelled "generations"; these groups are determined based on the writers' birth years, and each generation contains individuals born within a time span of 15 years. In the compilation of the groups, special attention was also paid to the authors' professions (each generation contains at

least one scientist), which may serve as a further useful tool in comparative studies (Petré et al. 2019: 88–90; 92–3; 100–4).

3.1.5 Selected authors from EMMA

Considering the impressive total wordcount of the corpus, I decided to reduce the data collection to the writing of just a few selected authors. An important question was that of how many would be needed for a representative overview of the language of the 17th century London–based elite.

Eventually, I chose four authors from the generations 2 and 3 of the corpus for the study. The writing that was analysed was that of Robert Boyle (1627–1691) and Margaret Cavendish (1623–1673) for generation 2, and Nathaniel Crouch (1640–1725) and William Salmon (1644–1713) for generation 3. While these writers all fulfilled the established criteria to be included in the corpus (see section 3.1.4) satisfactorily enough, adjustments had to be made for them to be eligible for this study. With the focus of the investigation lying on the use of relative pronouns in medical and/or scientific texts, the ideal author would have had to be of a relevant profession and/or to write predominantly in these domains (for instance, a doctor writing theoretical texts about medical treatments or a scientist documenting his experiments). Only Boyle fulfilled this requirement to a degree where his writing was exclusively of scientific nature. Numerous scientific texts were also available in the case of William Salmon, who worked as a doctor. However, for reasons of convenience (mainly to make the texts comparable in terms of size and year of production), his art manual *Polygraphice* (1672) was also included. The profiles of the four authors are displayed in Table 3.1:

generation 2	Year of birth/death	Profession(s)	Total wordcount in EMMA
Boyle, Robert	(1627–1691)	natural philosopher, chemist, physicist, inventor	2,082,984
Cavendish, Margaret	(1623–1673)	philosopher, poet, scientist, fiction–writer, playwright	1,393,983
generation 3			
Crouch, Nathaniel	(1640–1725)	printer, bookseller, historian	1,791,125
Salmon, William	(1644–1713)	doctor	2,889,362

Table 3.1: Authors from EMMA selected for the study; information taken from Petré et al. 2019: 89

A special case is that of Margaret Cavendish, Duchess of Newcastle–upon–Tyne. Her writing career lasted 15 years and was shorter than that of the other authors. Furthermore, she spent a large part of her life in exile in France and the Netherlands due to the Civil War, and only a limited amount of time in London. Therefore, her profile presents some differences to those of the other authors and was, generally, not a perfect match to the criteria established in the compilation. She was still included in the corpus as a member of a control group, useful to investigate the spread of linguistic changes throughout the London–based elite of the 17th century (Petré et al. 2019: 87–8; 110). Her texts were suitable for this particular study, given her background as a scientist (among other occupations) and her, still, considerable amount of writing. Furthermore, despite not being a special focus of this project, there was a possibility to see contrasts between her writing and that of the three male authors (the contrasts in the distribution of *which* and *the which* between genders in the Corpus of Early English Correspondence were addressed in Raumolin–Brunberg's (2000) study). Her language has been studied before; her variation in the use of third person singular *–s*/*–th* was examined in a small case study conducted by Petré et al. (2019: 110–111; see also section 4.2.2).

Another author who stands out in this group is Nathaniel Crouch. He grew up in a middle–class family of artisans and did, therefore, not have the same upper–class background as the other writers. He also did not attend university. Nevertheless, he was an important and successful publisher in London and had significant connections to the city and its elite (Fonteyn & Petré 2022: 98). As I also chose to address the authors' socioeconomic backgrounds in this project, such contrasts between them were especially interesting.

Analysing the full written material from each author would have been an undertaking far beyond the scope of this project, considering that their respective total wordcounts amount to between 1,393,983 and 2,889,362 tokens. Therefore, I decided to further reduce the number of texts to one per decade from each author. The texts were selected based on their length, to make them more practically comparable, and their year of production for an evenly distribution across the authors' writing careers. At this point, an inevitable but natural factor became clear: the authors did not all write for similar periods of time, given that they naturally did not all live to the same age and started their careers at different points during their life, producing more writing during certain decades than others. This was not considered a bigger issue, as the texts were still relatively evenly distributed throughout their careers. However, the numbers of texts varied, resulting in 4 texts from Boyle (composed between 1660 and 1691), while only 3 texts

from each of the other authors were analysed. The selected writing of Cavendish was composed between 1655 and 1668, Salmon's between 1672 and 1695, and Crouch's between 1678 and 1696. This resulted in a noticeable inequality, as it allowed me to analyse one additional decade of Boyle's writing. Also, in the case of Cavendish, the texts were produced in the years 1655, 1662 and 1668, resulting in a smaller temporal distance. Situations like these were, however, expected in a work centred around the authentic and full–length writing careers of real authors.

3.1.6 Data collection from EMMA

Once the texts had been selected, I generated wordlists and noted the total occurrences of the relative pronouns (the results are displayed in chapter 4). Subsequently, I collected the antecedents and classified them according to the same categories as those from MCEModESP. While the complete data from the 16th century was considered in the first corpus, the hits from the search string *which* and *the which* in the 17th–century texts had been reduced to samples of 100, randomly generated by AntConc (see section 3.1.2). The same method was adopted in the case of EMMA, as each text produced several hundred hits; numbers that were impossible to include in their entirety within the margin of this study. The other search strings, namely those for *the which*, *who*, *whom* and *whose*, did in most cases produce under 100 hits and I was, therefore, able to include all of them in the results (in the case of more hits, a random sample of 100 was generated here as well). When working with corpora, the data that is being analysed always constitutes a limited portion of actual language use in general. It is always just a sample that can be examined closely and that, in this case, is supposed to represent the larger text it stems from. The results and the related conclusions will, therefore, only be certain for the samples. However, they may lead to generalisations about the language used throughout the corpus (Barth & Schnell 2021: 137).

While a quantitative approach was taken in the first part of the study, in the case of the second corpus it was mostly qualitative, despite its specific design for the quantitative study of linguistic lifespan change (Petré et al. 2019: 85; see also section 3.1.7). This is due to the limited number of texts that was included, as a comprehensive and fully representative analysis of the authors' entire writing was not possible. Reaching complete representativeness is a virtually impossible undertaking in general, considering that not even extensive datasets such as EMMA can contain all linguistic choices made by a speaker community. The topic of representativeness in corpus studies should, therefore, be approached from a more general perspective and it should be kept in mind that text collections only represent a fragment (in the case of this study, only

small samples) of the real–world population of texts (Barth & Schnell 2021: 17; see also McEnery & Hardie 2011: 10–11).

3.1.7 Comparing the corpora

When comparing the two corpora used in the study, perhaps the most striking difference is their size. In terms of representativeness, it is generally true that the more wordform tokens a corpus contains, the better. After all, the smaller the portion of language, the fewer insights and less variation will it present in the use of a feature in context and potential patterns. It is also true, however, that the main criterion when it comes to evaluating the adequacy of a corpus for a given study is its suitability for the proposed research questions (Barth & Schnell 2021: 19; 25; 27). As discussed in the previous sections, both corpora seemed well–suited for this project. It is nevertheless necessary to point out their difference in size. With just over 1 million words, MCEModESP can be considered relatively small. This is not surprising, considering that it is a subsection of a specialised corpus. In comparison, EMMA presents far more material to analyse, comprising over 90 million words in the writing of 50 different authors (see Petré et al. 2019). Both corpora are static, and no texts are intended to be added, despite ongoing projects to improve EMMA's genre classification as well as implementing spelling normalisation (see Petré et al. 2019: 114 and Barth & Schnell 2021: 36–7).

As previously mentioned, the main prerequisite for the corpora to be included in this study was for them to contain Early Modern English scientific and/or medical texts, and to be suitable for diachronic research (covering, therefore, ideally the entire period or at least multiple decades of it to make the individual texts and their language comparable) (see McEnery & Hardie 2011: 94–5 and Barth & Schnell 2021: 37). This was achieved by both datasets, with MCEModESP covering the years 1500 to 1700, while EMMA focuses on the mid to late 17th century.

I briefly mentioned in section 3.1.6 that the data collection in the first half of the study was of quantitative nature, while in the second it was mostly qualitative. While counting, classifications and statistical methods lie at the base of quantitative research, in qualitative studies the focus does not lie on the frequency of the investigated feature (McEnery & Wilson 2001: 76). In MCEModESP, I counted all instances of *which*, *the which*, *who*, *whom* and *whose* that appeared in the subcorpus. Subsequently, I collected and categorised their antecedents. In the case of EMMA, however, only small samples from four authors in a 90–million–word

corpus were analysed. The focus, therefore, no longer lay on how frequently the relativisers appeared in the dataset, but how they were used in the selected texts. In McEnery's & Wilson's (2001: 76) words, in this case "the data [was] used only as a basis for identifying and describing aspects of usage in the language […]". Generally, both approaches are considered equally important to corpus linguistics, as both reveal important insights into the available data. Furthermore, they can often complement each other (McEnery & Wilson 2001: 76; McEnery & Hardie 2011: 2). In this study, for instance, the quantitative perspective provides a diachronic overview of how the relative pronouns developed in terms of their frequency. The following qualitative approach, on the other hand, makes it possible to study the contexts they appeared in and the factors that conditioned their use.

3.2 Statistical analysis

In sections 3.1 to 3.1.7, I discussed the corpora I used in the study and the data collection process, as well as the data classification. Calculating and comparing the frequencies of the relative pronouns throughout the datasets was an essential step to trace their diachronic development. Likewise, classifying them according to their antecedent revealed important insights into their preferred patterns and uses. However, the significance of numbers and percentages in tables has its limitations, if no additional operation is carried out. For this reason, it was necessary to test the obtained data for any statistical relevance in its distribution. In the next subsection, I will discuss the role of statistics in linguistic studies. In section 3.2.2, I will briefly introduce the software R and describe the statistical tests that were carried out with it in this study.

3.2.1 Statistics and linguistics

Statistics lies at the very base of scientific investigation. Without statistical analysis, it would be hardly possible to test theoretical hypotheses, or to assertively make any generalisations about the population a sample stems from. However, statistics has only recently begun to be recognised as an essential tool for linguistic studies and to spread from hybrid disciplines, such as psycholinguistics or computational linguistics, to the more traditional theoretical areas of research. While basic operations such as chi–squared tests have long been a regular sight in linguistic studies, more advanced statistical methods and analyses have become increasingly common only in recent years (Levshina 2015: 1).

The discipline of statistics can be further divided into descriptive and inferential statistics. The first can be used to describe the characteristics of a given sample, while the latter makes it possible to form conclusions about the general population based on the sample (Levshina 2015: 7). Both aims are very similar to what corpus linguistics is concerned with, i.e., studying representative samples to draw generalisations about the texts they stem from, and the language used in them (Barth & Schnell 2021: 17). Samples are used since it would be impossible to study a population in its entirety, and this is where statistical tests become crucial.

Statistical methods go beyond the counting and comparing of frequencies. The aim is to test theoretical hypotheses and to compare datasets with the expected results. Simultaneously, the absence of statistical significance can determine the next steps in a given study or the interpretation of the results. Additionally, statistical tests can indicate whether potential measurement errors may have occurred (Johannson 2017: 3–4).

As mentioned above, advanced statistical analyses have only recently found their way into linguistic studies. Their introduction is correlated with fast advancing, cross–disciplinary methods, but also with the increasing number of large–scale digital corpora. The latter require effective tools for data processing to make empirical evidence detectable. It is, however, true that hardly any existing software is suitable to tackle all characteristics and issues of natural language studies. The open–source programming language R is probably the most popular and frequently used one, as it shows more versatility than other tools and software (see section 3.2.2) (Gries 2017: 2–4).

In summary, statistics is a crucial part of linguistic studies to test theoretical hypotheses. It is a powerful tool that allows the researcher to compare findings to the expected results, indicating whether an assumption should be rejected or whether its validity is supported by the data. This being said, statistical tests can and should not be seen as ultimate proof for a hypothesis to be true, or entirely false. There may be situations where statistical relevance is not detectable, for example when the dataset is very small and/or not representative, or the wrong test was employed. In addition, even if a test shows evidence for statistical relevance, it will not provide answers as to why the data displays certain patterns. In other words, statistics is as an important tool to gain insights into a dataset, but its interpretation is, ultimately, up to the researcher (Levshina 2015: 15–16; Johansson 2017: 3–5).

3.2.2 Statistical analysis in the study

As described in sections 3.1.2 and 3.1.4, I collected and classified the data from the corpora according to an established set of criteria. After compiling tables that showcased the distribution of the relativisers (displayed in chapter 4), I calculated the resulting percentages. This step made it possible to directly compare the numbers of occurrences and to generate an overview of how the relative pronouns were used in the texts.

However, hypotheses cannot be proved through observation alone (Winter 2019: 5). Any conclusions based solely on the tables would not have been very reliable without further tests to support them. Statistical methods were needed to indicate whether the distribution of the data was in line with my hypotheses (see chapter 1), or whether they presented a different picture. In other words, a central question was whether the distribution of the relativisers was purely by random chance, or whether it likely happened according to a system. In the case of patterns, the question was whether they matched the expected distribution or whether it followed a different scheme.

I used R, one of the most popular tools for statistical studies in linguistics, to carry out the tests needed to further analyse my data. The programming language offers a broad range of functions and add–on packages that make it widely used in, especially, corpus–based and computational studies. Apart from methods for statistical analyses, effective tools for data retrieval, data manipulation and calculation, and result visualisation are also provided. Distributed under the GNU General Public License, R is freely available for download and may be used, studied, shared, and modified by its end users. The resulting R community is constantly working on new packages for specific tasks, and support is available from experts of all fields (Levshina 2015: 21; Gries 2017: 269).

The data was distributed in cross tables with several categories and variables, resulting in multiple rows and columns (see chapter 4). I imported the data into R to perform chi–squared tests, as the software has a special function to run them in cross tables. The aim of these tests was to obtain the *p*–value of each distribution, to gain insight into whether they likely occurred according to a pattern or not.

In my data, the maximum value established for the *p*–value was 0.05[10]. It is important to remember that the *p*–value should not be treated as proof that hypotheses are true or false, or as an indicator of relevance. It is merely a statement about the probability of falsely rejecting the null hypothesis[11]. With cross tables, the null hypothesis is that the rows and columns are independent of each other and do not reveal any new information about the data (Johansson 2017: 26; 30). In the context of this study, the null hypothesis would be that there is no relation between the relative pronouns and the type of antecedents they occur with. Their distribution is by random chance and the authors did not use the relativisers according to any actual system. It goes without saying that, in reality, it is assumed that there may very likely be patterns, and the ideal outcome of this project is to identify and describe them. However, the probability of falsely rejecting the null hypothesis should, of course, be as minimal as possible.

Subsequently, I generated association graphs of the distributions with the help of R's assocplot and assoc functions (the latter is included in the Visualizing Categorial Data (VCD) package). The resulting graphs, as well as the following operations, are illustrated in section 4.3.

An important question concerns what the results from the statistical analysis mean in practice. In Johansson's words, "statistical significance does not [automatically] mean relevant, or interesting or even easy to detect" (2017: 7). Considering the numerous previous attempts at defining the distribution of *which* and *the which* (see sections 2.1.4 and 2.1.5), the results from this study are not expected to unequivocally confirm or reject existing conclusions. The statistical test that was carried out and the resulting *p*–values should mainly be seen as helping indicators, meant to provide clues on potential patterns.

Another crucial factor that needs to be kept in mind is the general scope of this project and the, relatively, limited amount of data that was analysed. Even in cases where the *p*–value did not indicate any statistical relevance, it cannot be excluded that with a larger dataset significance would have been observable. The way I further tested this with the help of association graphs will be explained in section 4.3.

[10] It is generally agreed among the scientific community that only *p*–values below 0.05 are considered sufficient evidence against the null hypothesis. Depending on the study, the threshold may also be set at 0.01 or 0.001 (Johansson 2017: 47; Winter 2019: 168).

[11] In statistics, the null hypothesis states that there is no difference between different groups, or association between variables, etc. The aim of the research (or alternative) hypothesis is to discard this claim (Levshina 2015: 8–9; Johansson 2017: 6; Winter 2019: 158).

Related to the extent of my data is the size of the individual samples. As described in sections 3.1.2 and 3.1.4, I only analysed a randomly generated sample of 100 in the case of *which*. The occurrences of *the which*, *who*, *whose* and *whom* were, in most cases, considered in their entirety. As a consequence, the statistical cross table test contained limited samples against total numbers. This needs to be taken into consideration, underlining once again that any conclusions are only certain for their specific sample of data. Such a situation was, however, expected: it was simply not possible to include all instances of *which* in the study, and the generated samples serve a representative function of its distribution throughout the texts they stem from.

3.3 Summary

This chapter's focus was the methodology employed in my study. First, I explained my choice of using Early Modern English scientific prose, as this text genre seemed particularly likely to feature instances of *the which*. Subsequently, in sections 3.1.1 and 3.1.2, I introduced the first dataset, namely MCEModESP. This corpus proved to be well–suited for the project in terms of the covered timespan and genres, but lacked metadata about its texts and authors. For this reason, it was necessary to add a second corpus to my study. The collection and analysis of the data from EMMA was discussed in sections 3.1.4 and 3.1.6. Compared to the analysis of the first dataset, I took a more qualitative approach in this half of the project, as I investigated selected writing of four individual authors.

I concluded the chapter with section 3.2, where I introduced R and its relevance for linguistic studies. To further analyse my data, it was necessary to carry out tests to confirm statistical relevance and the presence of patterns.

4. Results and discussion

In this chapter, I am going to present and discuss the results of my study. In section 4.1, I will illustrate the total frequencies of *which*, *the which*, *who*, *whom* and *whose* in MCEModESP. In addition, their distribution among the different types of antecedents will be displayed. In the following section, 4.2, the focus shifts to EMMA and the way the relativisers were used across the texts that I analysed. One subsection will be dedicated to each of the four authors. Finally, in section 4.3, I will discuss the statistical significance of the distributions. All results will be illustrated with the help of tables and graphs, as well as examples to facilitate the discussion.

4.1 Total frequencies of the relativisers in MCEModESP

I took a quantitative approach to analyse the data from MCEModESP. Further, I divided the corpus into two unequal halves, namely texts composed in the 16th century and writing from the 17th century. Considering that the instances were far more numerous in the first half than in the second, I calculated the corresponding percentages of the relative pronouns throughout data. The total frequencies of the relativisers are displayed in Table 4.1:

Relative pronoun	**Total occurrences in MCEModESP (16th century)**		**Total occurrences in MCEModESP (17th century)**	
	N	%	N	%
which	129	~79.6	1418	~83.3
the which	29	~17.9	188	~11
that which	0	0	6	~0.3
who	1	~0.6	48	~2.8
whom	1	~0.6	21	~1.2
whose	2	~1.2	21	~1.2

Table 4.1: Total occurrences of the relative pronouns which*,* the which*,* (that which)*,* who*,* whom *and* whose *in MCEModESP*

Besides the investigated pronouns, the relativiser *that which* is also featured in Table 4.1. No explicit search string for this relative pronoun had been conducted, but as its occurrences (albeit limited in number) stood out during the data collection and later analysis, I decided to include it. In Middle English, it had been possible to freely combine *that* with other relativisers, and elements such as *that which* or *who that* were common (Dekeyser 1984: 61; see also section 2.1.3.1). Generally, it would seem like these combined pronouns are rather understudied, as they are rarely included in the literature. However, considering the occurrences in my data and other accounts (see, e.g., Fischer 1992: 302–3; Brinton & Arnovick 2006: 334), these relativisers may still have been common in Early Modern English. Two instances of *that which* are illustrated in examples [15] and [16] (marked in italics) from the 17th–century–half of my data; no combined pronouns (apart from *the which*) appeared in the 16th–century texts.

[15] Take sal inter *that which* is white like glass, […] (Philip Stanhope, mid–17th century. *MS Wellcome 762*. MCEModESP)

[16] Take A Quart of Cream and boil therein A good handful of the Inner Bark of Elder, *that which* is green and boil them together till it comes to An Oil […] (Elizabeth Jacob (and other authors), between 1654 and c.1685. *MS Wellcome 3009*. MCEModESP)

When looking at the total frequencies of *which* and *the which*, it is not surprising that *which* occurred far more frequently than the combined form. As pointed out by Nevalainen et al. (2011: 10), *the which* was always the minority variant. Their claim contradicts Mustanoja's (1960: 199) conclusion that it was generally preferred over plain *which* in 15th–century prose (see also section 2.1.5). At least in this dataset, *the which* was undoubtedly the less popular choice.

The decline of *the which* can also be observed in the data. In fact, while its percentage of occurrence in the 16th–century texts was ~17.9%, in the data from the 17th century it was ~11%. This development was expected, as the use of *the which* declined rapidly after 1600 (see, e.g., Rissanen 1999: 297; see also section 2.1.4).

The development of the animacy distinction, which led to (*the*) *which* gradually being replaced by *who* with human antecedents, reached stability in the second half of the Early Modern English period. In the early 17th century, however, it was still possible to use both with personal referents (see, e.g., Rissanen 1999: 294; see also section 2.1.3.3). The following two

examples show a divine [17] and a personal [18] antecedent (underlined, while the relativiser is marked in italics):

[17] And this have I written in the praise of almighty god *which* gave me grace so to devise for my health and prolonging my life. (Unknown author, between 1499 and 1599. *MS Hunter 135*. MCEModESP)

[18] Such an humour is common to maids, *which* lead a sitting, idle and ill ordered life, […] (Jean Liébault, 17th century. *MS Hunter 303*. MCEModESP)

The occurrences of relative *who* and its oblique counterpart *whom* are, generally, rather limited in MCEModESP. Only one instance of each appeared in the 16th–century data, with only a slight increase in the following century in terms of the percentages covered. The few appearances of *who* and its inflected forms may be traced back to the text type. In the case of recipes, for example, bullet point–like listings and instructions may be more prominent than elaborate literary passages. Consequently, these texts may rather describe ingredients and substances than people, and perhaps contain fewer subordinate clauses altogether (see Taavitsainen et al. 2011: 23–4). More qualitative insights into the individual texts would be needed to confirm this.

Possessive *whose*, on the other hand, can refer to both personal and nonpersonal antecedents, given that neither *which* nor *that* (the most popular relativiser in Early Modern English) has a possessive form (Dekeyser 1984: 61; Rissanen 1999: 294). No dramatic changes were, therefore, expected in its frequency. Still, it was almost absent from the 16th–century data.

The Pearson's chi–squared test that I carried out with R revealed a *p*–value of 0.07547. The value was above the established threshold of 0.05 and did, therefore, not indicate statistical significance in the diachronic development between the two parts of the corpus.

4.1.1 Distribution of the relativisers in MCEModESP

In this section, I will take a closer look at the grammatical profiles of *which* and *the which* in MCEModESP. The total frequencies of the relativisers, including also *that which*, *who*, *whom* and *whose*, which were illustrated in the previous section, revealed important insights into their diachronic development. Overall, it would seem like the observed instances confirm

previous conclusions. *The which* still appeared in the data from the 17th century, but the instances of plain *which* were far more numerous. The use of the combined pronoun also decreased over the period. Of course, larger datasets would be necessary to state this with confidence, as especially the texts from the 16th century only counted a limited number of tokens. This being said, my first research question addressed the actual patterns governing the distribution of the pronouns:

> *How are the relativisers* which *and* the which *used in Early Modern English scientific/medical prose and to what extent do they act as free variants? How are they distributed in terms of antecedents and are there any patterns? If yes, which ones confirm previous conclusions and which ones do not?*

A central aim of my study was, therefore, to explore the antecedents *which* and *the which* occurred with, attempting to identify patterns that may hint at subtle differences in their use. An underlying goal was to reject the assumption that the relativisers functioned as free variants in all contexts and were fully interchangeable; a research question that remained largely unanswered in Raumolin–Brunberg's (2000) study.

The antecedents the relative pronouns occurred with in the corpus are displayed in Table 4.2. The 17th–century half represents a randomly generated sample of 100 (see section 3.1.2). In the 16th–century half, the occurrences of *which* and *the which* were included in their entirety:

	MCEModESP (16th century)		**MCEModESP (17th century)**		
	which	*the which*	*which*	*the which*	*that which*
Personal reference (without pronouns)	1	0	0	1	0
Personal pronouns	2	0	0	0	0
Deity	1	0	0	0	0
Nonpersonal reference (without pronouns)	100	28	85	87	5
Nonpersonal pronouns	5	1	12	2	0
Total	*109*	*29*	*97*	*90*	*5*
Restrictive clause	47	16	24	24	3
Non–restrictive clause	62	13	73	66	2

Total	*109*	*29*	*97*	*90*	*5*
Subject	82	18	84	33	4
Object	17	0	12	7	1
Prepositional phrase	10	11	1	50	0
Total	*109*	*29*	*97*	*90*	*5*
Plural antecedent	19	4	8	35	0
Singular antecedent	74	22	76	48	5
Sentential antecedent	16	3	13	7	0
Total	*109*	*29*	*97*	*90*	*5*
Determiner	19	10	0	10	1
Others	1	1	2	0	0
Interrogative pronoun	0	0	1	0	0

Table 2.2: Grammatical profiles of which *and* the which *(and* that which*) in MCEModESP. Design taken and adapted from Raumolin–Brunberg (Table 3; 2000: 212)*

The design of the table, including its categories and variables, closely follows that proposed by Raumolin–Brunberg (2000: 212) in her study. The first section focuses on the animacy distinction. Here, I included categories for personal and nonpersonal nouns, as well as the Deity. In addition, a distinction was also made between personal and nonpersonal pronouns. As becomes evident when looking at the table, personal *which* and *the which* were very rare in my data. While in the texts from the 16th century there were four instances of personal reference (including the Deity), only one occurrence of *the which* appeared with a personal noun phrase in the 17th–century texts. Generally, a shift was expected, as (*the*) *which* eventually became confined to nonpersonal antecedents during the second half of the Early Modern English period. However, it is interesting how early the authors in MCEModESP seem to have complied with the animacy distinction, as, in general, personal (*the*) *which* was still common in the early 17th century (Rissanen 1999: 294). A reason for this may be the high prestige of medical writing that may have required an earlier compliance with conventions of politeness. Other standardisation processes, such as the fixation of spelling were, however, slower in this genre compared to others (Taavitsainen 2000: 147).

Dekeyser (1984: 69) stated in his study that *who* was mainly used in non–restrictive clauses and *that* in restrictive ones, while *which* had a more neutral role (see also section 2.1.3.2). This is not reflected in the next section of Table 4.2, where *which* is clearly preferred in non–restrictive clauses. The same is true for *the which* in the 17th–century data. The *p*–values that I calculated did, however, not indicate any statistical significance in this distribution.

Statistical relevance (*p*–value=0.000157 in the 16th–century texts and *p*–value=3.724e–15 in the 17th–century data) was found in the following section, displaying the distribution among subjects, objects, and prepositional phrases. In the texts from the 16th century, *which* occurred far more frequently in subject position than in any other function. *The which*, on the other hand, did not appear a single time as the object, but occurred both as the subject and in prepositional phrases. The situation is similar in the 17th–century data, except for *the which* occurring for the greater part in prepositional phrases. This is in line with Raumolin–Brunberg's (2000: 221) observed development. Her data, covering the Late Middle English period, indicated that *the which* was eventually preferred with prepositions, as in example [19] (underlined; *the which* is marked in italics), while *which* was favoured as the subject. My data, containing texts from the following centuries, reflects this trend and shows that it may likely have continued in the last decades where *the which* was in use.

[19] Take three ounces of this wine, <u>with</u> *the which* you shall mix a drachm and a half of citron myrobalan made in powder that is fine and small, […] (Jean Liébault, 17th century. *MS Hunter 303*. MCEModESP)

The next section of the table shows how both *which* and *the which* were mainly used with singular antecedents in MCEModESP, except for a clear rise of plural *the which* in the 17th century. What was unexpected are the few occurrences of *the which* with sentential antecedents. According to Fischer (1992: 303–4), the pronoun was mostly used in non–restrictive, continuative relative clauses, as this clause type is rather common in academic prose. This is, however, not reflected in my data and sentential antecedents (as underlined in example [20]) are, even, the ones with the least occurrences in both centuries. The *p*–value calculated for the 17th–century data in this distribution was 1.341e–05 and, therefore, indicated high statistical significance.

[20] If proud flesh grow in a wound for *the which* many Surgeons do use alum calcined or powder of chaperons, ... (Unknown author, between 1499 and 1599. *MS Hunter 135*. MCEModESP)

Finally, the last section of Table 4.2 features instances that can be classified as, respectively, determiners, other functions, and interrogative pronouns[12]. Originally, I had also included a category for unclear cases, but since I did not encounter any in the analysed samples from MCEModESP or EMMA the category is not reported in the tables. In the four instances labelled as "others" the pronouns appeared in free relative clauses. In the 16th–century data, both *which* and *the which* frequently occurred as determiners. In the 17th–century texts, however, only *the which* appeared in this function. This is illustrated in examples [21] and [22], where the referents are underlined:

[21] […] take the patient by the Same Arm and lift his arm upright through *which* lifting it shall bolt into joint, […] (Unknown author, between 1499 and 1599. *MS Hunter 135*. MCEModESP)

[22] Of *the which* past you may form pills. (Christopher White (Junior), between 1679 and 1755. *MS Hunter 43*. MCEModESP)

As mentioned above, the design of the table and its categories closely follow those suggested by Raumolin–Brunberg (Table 3, 2000: 212). Since she analysed data from the Late Middle English period, the numbers are difficult to compare directly. However, emerging trends that she spotted seem to be confirmed in my data, which covers the following two centuries. For instance, she observed that *which* was increasingly preferred as the subject in relative clauses, while *the which* was common with prepositions. This is also the case in MCEModESP, as well as the fact that both pronouns occurred more frequently with singular than with plural antecedents.

[12] In English, interrogative pronouns are identical in form to relative pronouns, but introduce either direct or indirect questions instead of relative clauses. The animacy distinction applies here as well, and *who* is reserved for human subjects (Peters 2013). Interrogative (*the*) *which* is, strictly speaking, a determiner and not a pronoun (Payne & Huddleston 2002: 428).

4.1.2 Sentences with multiple relativisers

A special distribution of the relative pronouns stood out in the data from this corpus: the dataset contained a number of instances where different relativisers were used in the same sentence. These sentences, hence, featured several relative clauses and were overall rather long and complex.

The trend towards longer and more embedded sentences had started in Middle English and continued in Early Modern English (van Gelderen 2006: 171). As a consequence, particularly in 16th–century prose the complexity of subordinating structures increased, following Latin models (Franz 1939: 427–73; cited in Görlach 1991: 122). Especially formal registers, including legal discourse, often contained long sentences with several subordinate clauses (see, e.g., Lehto 2012; 2015). Such structures would hardly be expected in spoken language. Three such examples are [23], [24] and [25], where the pronouns are marked in italics:

[23] Take pearls *that* are Indifferent great, *which* are not round, *which* you shall have at the jewellers very reasonable, […] (Elizabeth Jacob (and other authors), between 1654 and c.1685. *MS Wellcome 3009*. MCEModESP)

[24] […], use roasted meats, such as are hens, kid, veal; mutton, partridges, and such like, *which* are not easily corrupted, *the which* shall be dressed with odoriferous spices, […] (Jean Liébault, 17th century. *MS Hunter 303*. MCEModESP)

[25] […], out of this dialatation is bread that coat *which* We commonly Reticularis, *the which* indeed as Galen says ought no ways to bear that name of a coat, […] (John Browne, between 1599 and 1699. *MS Hunter 92*. MCEModESP)

In example [23], the alternation between *that* and *which* may be justified by the author's wish to avoid redundancy. However, *which* still occurs twice in consecutive clauses, resulting in repetition.

The situation is different in examples [24] and [25]. Both sentences are rather long, which might account for the need of different relativisers, to avoid repetition and ambiguity. In addition, *which* and *the which* seem to perform slightly different functions in these examples. Traditionally, the uses of the two pronouns are considered virtually identical, raising the question of whether they acted as free variants (Raumolin–Brunberg 2000: 221). In these cases,

however, their roles appear slightly different. In example [24], *which* refers to *roasted meats*, a general collective term for the elements the author listed right before. Once these items have been defined as *hens*, *kid*, *veal*, etc., they are referred to using *the which*. It seems, thus, like the combined pronoun is more suitable for already identified and described antecedents, while *which* can also be used for general, undefined referents. Support for the latter is found in previous studies. For example, despite focusing on the distribution of *which* in opposition to *who*, Sato (2019: 277) stated that Shakespeare clearly preferred *which* when referring to collective nouns.

The same construction can be observed in example [25]. Plain *which* refers to the antecedent *coat*, which is then further defined by its name *Reticularis* in a restrictive relative clause. Once the *coat* has been labelled, there is no ambiguity about its identity, and *the which* is used in the following clause.

These different functions are highly speculative. It is not a pattern that is typically pointed out in the literature. The only related characteristic of *the which* is that it was preferred in non–restrictive continuative clauses. This clause type is often found in academic and scientific prose, as long embedded sentences with several subordinate clauses are common here. In these cases, the combined relativiser was often considered a clearer link to the antecedent, while plain *which* or *that* may have caused ambiguity (Fischer 1992: 303–4). Such sentences may, likely, also require different pronouns in the individual clauses to avoid repetition.

This construction appeared several times throughout my data, in texts written by different authors. In MCEModESP, I counted a total of 22 sentences which contained different relative pronouns. As these structures were not, initially, a focus of my data collection, it is possible that the corpus contains more sentences, which were overlooked. Out of the observed instances, 3 featured both *which* and *the which*. In these cases, *which* always preceded *the which*, and never the other way around. This hints at a more or less fixed pattern that served a specific purpose. Sentences featuring *which* and *that* were, however, far more numerous. Their functions appeared very similar and, likely, hint at a stylistic choice to avoid repetition.

There were also instances where *that* appeared in the same sentence as *the which*. In these cases, too, *that* appeared first and in a restrictive relative clause, and was replaced by *the which* in the following clause when the referent had been identified. In example [26], all three relativisers (marked in italics) appear in the same sentence:

[26] […], corrupt the meats and engender many crudities (*which* are the cause of infinity diseases) through the floating and disordered moving of the meats *that* is made in the stomach: *the which* after meal have need of rest not of stirring to be digested, […] (Jean Liébault, 17th century. *MS Hunter 303*. MCEModESP)

Overall, these examples do not necessarily indicate a distinction between the roles of *which* and *the which* in particular. Relative *that* was also common in these sentences and performed similar functions to *which*. Furthermore, it is not certain whether the first clause is necessarily always restrictive in this construction. It may simply be that plain *which* provides more inherent information about the referent, and that it needs to immediately succeed the main clause to avoid ambiguity, especially when there are multiple relative clauses. Rather, the observed instances present a pattern that seems exclusive to *the which*: when referring to the same antecedent, the combined pronoun appears in a non–restrictive clause and never precedes the other relativiser.

4.1.3 Summarising the results from MCEModESP

In the last three sections, I discussed the results from the first dataset employed in my study. Analysing the data from MCEModESP, I took a quantitative approach and investigated the total frequencies of *which*, *the which*, as well as *who* and its inflected forms throughout the corpus. This step, just like an analysis of the grammatical profiles of *which* and *the which*, were necessary to provide answers to the research questions 1 and 2, as well as their corresponding hypotheses (see chapter 1).

My first research question concerned the distribution of the pronouns in scientific writing. Generally, the authors in MCEModESP seemed to mostly adhere with the standardisation processes and conventions that were established in Early Modern English. For example, I had hypothesised that personal (*the*) *which* would still be common at the beginning of the period. However, already the 16th–century data featured only very few such instances. A further decline can be observed in the following century; a development that was expected.

Concerning the frequencies of *which* and *the which*, the plain relativiser was indeed far more present in the writing than the combined pronoun. Also, I had hypothesised that the use of *the which* would further diminish during in the 17th century–data. As Table 3 showed, its percentage among the investigated relativisers declined from 17.9% in the 1500s to 11% in the

1600s. Unfortunately, a more detailed overview of its diachronic development cannot be provided, due to the nature of MCEModESP. The dataset's metadata mostly indicates approximated dates of composition (e.g., between 1599 and 1699), which makes it difficult to say whether *the which* still appeared at the end of the period. This, in turn, did not allow me to draw conclusions on whether the scientific texts in this corpus were particularly innovative or conservative. Instead, this question will be readdressed in sections 4.2.1 to 4.2.5, where I will discuss the results from EMMA and its more precisely dated texts.

In hypothesis 1b, I also addressed the use of the pronouns. *Which* was prominent as the subject in relative clauses, as also observed by Raumolin–Brunberg (2000: 221). She, further, noticed in her data that *the which* was preferred in prepositional phrases, especially around 1500. This trend seems to be partially confirmed in my results. The combined relativiser appeared mostly with prepositions in my 17th–century data, but it was also popular as the subject in the previous century. Fischer's (1992: 303–4) claim that *the which* was mainly used with sentential antecedents was, instead, not confirmed in this corpus.

In terms of restrictiveness, both relativisers were clearly preferred in non–restrictive clauses (except for *the which* in the 16th–century texts, where it appeared more frequently in restrictive ones). While this was expected with the combined pronoun, Dekeyser's (1984: 69) claim that plain *which* was 'neutral' in terms of restrictiveness was not confirmed.

The, perhaps, most complex concern of this study was the question of whether *which* and *the which* acted as free variants. Raumolin–Brunberg (2000) did not reach any conclusions on this matter in her study. Other scholars (e.g., Dekeyser 1984 or Nevalainen & Raumolin–Brunberg 2002) treat the pronouns as one and the same variable, therefore advocating for interchangeability. Generally, I did not spot any patterns in my data that would indicate striking distinctions in their use. What, however, stood out were those instances where the two relativisers were used within the same sentence (or sequence of sentences). I observed several structures where *which* appeared in a restrictive clause and, when referring to the same antecedent, *the which* would follow in a non–restrictive one. In these cases, it seems like the pronouns carried slightly different roles related to the information parameter. This being said, this pattern reveals more about specific uses of *the which* than about the relation between the relativisers. In other instances, in fact, the role of plain *which* was performed by *that*, while *the which* maintained its non–restrictive function and succeeded the first relative clause.

As far as my second hypothesis is concerned, it is difficult to state with certainty whether the texts displayed more variation at the beginning of the period than towards the end. Once

again, this is due to the vague dating of the manuscripts. As mentioned above, the animacy distinction was already well implemented in the 16th–century half of MCEModESP. Similarly, there was indeed a decline in the use of *the which*. It is necessary to look at the results from the second corpus, EMMA, to further confirm this hypothesis and any previous claims.

4.2 *Which* and *the which* in EMMA

After discussing the data from MCEModESP, I will now proceed to illustrating the results from EMMA. Each of the four authors in the study presented rather different profiles and uses of the relativisers. For this reason, an entire subsection will be dedicated to each of them. Just like in sections 4.1 and 4.1.1, the results will be displayed in tables and examples will be provided.

4.2.1 Frequencies and distribution of the relativisers in Robert Boyle's texts

The first author featured in EMMA, whose writing I analysed, was Robert Boyle (1627–1691). During his lifetime, Boyle worked as a natural philosopher, chemist, physicist, and inventor, and wrote predominantly in the domains of medicine, chemistry, and physics (see Table 3.1 and Petré et al. 2019: 89). His overall profile, thus, fit the focus of this study extremely well. The main goal of this project had been to investigate the distribution of *which* and *the which* in academic prose (focussing mainly on scientific and medical texts), and Boyle was one of the most suitable candidates included in EMMA. Furthermore, his total wordcount of 2,082,984 was more or less evenly distributed over a three–decade long career.

A native Irish man, Boyle is regarded as one of the most influential figures of the Scientific Revolution (Anstey 2000: I; MacIntosh 2006: 4). There is no full–scale biography of his life, but as the son of the 1st Earl and Countess of Cork, it can be safely assumed that he enjoyed a privileged background and had access to higher education. Particularly interesting about his philosophy is his approach to the experimental sciences, which he combined with his theological interests (MacIntosh 2006: 3–4). It is, thus, not surprising that his works featured in EMMA fall into the categories of science and, sometimes, also religious prose (only scientific texts were included in the analysis).

Given the extent of his writing, as well as the fact that texts from 1660 to the early 1690s were available, four of his works (one per decade) were included in the study. In the case of the

other authors, I only included three, as their careers were shorter. The total frequencies of the relativisers in Boyle's writing are displayed in Table 4.3:

Relative pronoun	**Boyle Text 1 (1660)[13]**		**Boyle Text 2 (1672)[14]**		**Boyle Text 3 (1682)[15]**		**Boyle Text 4 (1691)[16]**	
	N	%	N	%	N	%	N	%
which	617	~81.7	444	~83.6	326	~84	489	~86
the which	0	0	0	0	0	0	0	0
who	56	~7.4	18	~3.4	20	~5.1	39	~6.8
whom	11	~1.5	7	~1.3	15	~3.9	9	~1.6
whose	71	~9.4	62	~11.7	27	~7	32	~5.6

Table 4.3: Total occurrences of the relative pronouns which*,* the which*,* who*,* whom *and* whose *in Robert Boyle's selected texts*

What immediately stands out is the fact that Boyle did not include *the which* in any of the four analysed texts. Not even in his earliest text from 1660 did he use the combined form and, instead, preferred other relative pronouns such as *which* and, presumably, *that*. Considering that *the which* had generally disappeared from writing by 1650 (Rissanen 1999: 297; see also section 2.1.4), Boyle's writing can be said to follow the timeline observed in previous studies. Despite the more conservative nature of scientific writing (Taavitsainen 2000: 147), this author may have considered *the which* an outdated option.

It is important to note that *which* was never used with human antecedents in these texts, indicating that the animacy distinction was well–underway in the second half of the 17th

[13] *New Experiments Physico–Mechanicall, Touching the Spring of the Air, and its Effects (Made, for the Most Part, in a New Pneumatical Engine): Written by Way of Letter to the Right Honorable Charles, Lord Vicount of Dungarvan, Eldest Son to the Earl of Corke/by the Honorable Robert Boyle, Esq.* (Robert Boyle, 1660)

[14] *Tracts Wtten by the Honourable Robert Boyle Containing New Experiments, Touching the Relation Betwixt Flame and Air, and about Explosions, an Hydrostatical Discourse Occasion'd by some Objections of Dr. Henry More against some Explications of New Experiments Made by the Author of these Tracts: to which is Annex't, An Hydrostatical Letter, Dilucidating an Experiment about a Way of Weighing Water in Water, New Experiments, of the Positive or Rlative Levity of Bodies under Water, of the Air's Spring on Bodies under Water, about the Differing Pressure of Heavy Solids and Fluids.* (Robert Boyle, 1672)

[15] *New Experiments Physico–Mechanical, Touching the Air*. (Robert Boyle, 1682)

[16] *The General History of the Air Designed and Begun by the Honble. Robert Boyle...* (Robert Boyle, 1691)

century, or that at least Boyle seemed to strictly adhere to it (see Table 4.4 below). Instead, *who* and *whom* were used to refer to personal or divine antecedents. In a few isolated cases, I observed instances of personification, as in examples [27] and [28] (the antecedents are underlined):

[27] Their Horſes Hair ſtood upright, like Briſtles, with the vehement Cold, *who* ſtood ſhaking, and refuſed to eat any thing till they came down. (Boyle, 1691. EMMA)

[28] But as for the Lungs themselves, *who* want Fibres to distend them, they may fitly enough be compar'd to a Bladder; … (Boyle, 1660. EMMA)

Personification was not uncommon in Early Modern English literary texts (for instance, it is frequently found in Shakespeare's writing), and can be traced back to notional gender distinctions (Abbott 1966 [1870]: 179–80; Franz 1924: 295; both cited in Johansson 2012: 784; Nevalainen & Raumolin–Brunberg 1994: 182–4). In example [27], *who* refers to an animate antecedent, albeit not a human one. Notional gender associations with inanimate elements, such as body parts, were observed by Nevalainen & Raumolin–Brunberg (1994) in the data from their study. Investigating the rise of the neuter possessive pronoun *its*, they noted that *he*/*his* was the preferred pronoun for certain inanimate nouns, while *it*/*its* was preferred for others. In addition, they observed variation and change over the period (183–4). Keeping this in mind, it is not surprising that my data contained instances where human body parts were personified, as in example [28], where *who* refers to *the Lungs themselves*. Redirecting the focus to *which* and its antecedents, Table 4.4 showcases its distribution in four randomly generated samples of 100 in Boyle's texts:

	Boyle Text 1 (1660)	**Boyle Text 2 (1672)**	**Boyle Text 3 (1682)**	**Boyle Text 4 (1691)**
	which	*which*	*which*	*which*
Personal reference (without pronouns)	0	0	0	0
Personal pronouns	0	0	0	0
Deity	0	0	0	0
Nonpersonal reference (without pronouns)	90	86	85	82

Nonpersonal pronouns	10	12	15	13
Total	*100*	*98*	*100*	*95*
Restrictive clause	42	27	56	19
Non–restrictive clause	58	71	44	76
Total	*100*	*98*	*100*	*95*
Subject	22	11	25	39
Object	36	56	35	31
Prepositional phrase	42	31	40	25
Total	*100*	*98*	*100*	*95*
Plural antecedent	12	17	25	20
Singular antecedent	74	61	57	60
Sentential antecedent	14	20	18	15
Total	*100*	*98*	*100*	*95*
Determiner	0	0	0	0
Others	0	2	0	4
Interrogative pronoun	0	0	0	1

Table 4.4: Grammatical profile of which *in Robert Boyle's selected texts. Design taken and adapted from Raumolin–Brunberg (Table 3; 2000: 212)*

As mentioned above, no instances of *which* were found with human or divine antecedents. The relativiser was used in both restrictive and non–restrictive relative clauses, with non–restrictive ones predominating in the texts from 1672 and 1691. Only in those two texts was the situation similar to MCEModESP, where *which* was clearly preferred in non–restrictive clauses.

The next section of the table shows that *which* was used throughout all texts as the subject, object or in prepositional phrases. The distribution here is relatively even and does not show any particular changes over time, apart from the 1672–text featuring the most instances of *which* in object position and the least as the subject. This is a striking difference compared to MCEModESP, where the authors mostly used *which* in subject position.

Similarly, the next section presents a very different picture from the first dataset. It appears that Boyle used *which* far more frequently with plural antecedents than with singular or sentential ones. In MCEModESP, *which* had been preferred with singular referents.

Finally, the instances labelled as "others" refer to free relative clauses, as well as adjectival, numeral, or adverbial antecedents. Overall, it would seem like Boyle used the relativiser differently from most authors in the first dataset. However, his texts also contained instances where *that* and *which* appeared in the same sentence, as in example [29]:

[29] THe thing *that* is wont to be admired, and *which* may pass for our second Experiment is this, […] (Boyle, 1660. EMMA)

In this case, it would seem like the two pronouns (marked in italics) were used to avoid redundancy.

4.2.2 Frequencies and distribution of the relativisers in Margaret Cavendish's texts

The second author featured in this study is Margaret Cavendish, Duchess of Newcastle–upon–Tyne (1623–1673). As mentioned in section 3.1.5, she presented a special case not only in this project, but also in the compilation of EMMA. In fact, one of the main criteria for inclusion in the corpus, namely a demonstrable link to the London society, only partially applied to her. In 1645, she married William Cavendish, Duke of Newcastle, who had left England when Cromwell seized power during the Civil War. The couple married in Paris and spent the next fifteen years in exile in France and the Netherlands. Apart from occasional business matters, carried out through correspondence, their ties with London were rather loose during this period. Furthermore, the main exposure to English they received must have been each other. As a consequence, it can be assumed that Cavendish's language must, at least to some degree, have been influenced by that of her husband and vice versa. It should also be noted that her husband was 31 years older and may have, altogether, used different linguistic features (Petré et al. 2019: 110–11).

This influence has been the focus of a past case study, carried out by Petré et al. (2019) who investigated her use of third person singular *–s*/*–th*. Their data, also taken from EMMA, revealed a striking rise of the *–th* ending during the period she spent in exile. Their natural conclusion was that her husband's language may have been the main driving force behind this

switch, as he belonged to a generation that still generally favoured *–th* and this ending also prevailed in his writing. Once their time in exile had ended and the couple had returned to London, Cavendish quickly readopted the *–s* ending and only sporadically used *–th* after the mid–1660s (110–11).

The study showcased the influence social networks can have on an individual's language. As Cavendish, more than likely, accommodated her use of the third person singular verbal endings to that of her husband, there are grounds to believe that this may have been the case with other linguistic features as well. For instance, her use of the relative pronouns may have changed during the years spent in exile. Before turning to the antecedents *which* and *the which* occurred with in her writing, the total occurrences of the relativisers are displayed in Table 4.5:

Relative pronoun	**Cavendish Text 1 (1655)[17]**		**Cavendish Text 2 (1662)[18]**		**Cavendish Text 3 (1668)[19]**	
	N	%	N	%	N	%
which	728	~96.3	502	~82	361	~93.3
the which	0	0	2	~0.3	0	0
who	21	~2.8	67	~10.9	12	~3.1
whom	2	~0.3	19	~3.1	6	~1.5
whose	5	~0.7	22	~3.6	8	~2.1

Table 4.5: Total occurrences of the relative pronouns which*,* the which*,* who*,* whom *and* whose *in Margaret Cavendish's selected texts*

Despite appearing only two times, in the text from 1662, *the which* is present in Cavendish's writing. The two occurrences (in italics) are displayed in examples [30] and [31]:

[17] *The Philosphical and Physical Opinions Written by Her Excellency the Lady Marchionesse of Newcastle.* (Margaret Cavendish, 1655)

[18] *Orations of Divers Sorts Accommodated to Divers Places Written by the Lady Marchioness of Newcastle.* (Margaret Cavendish, 1662)

[19] *Ground of Natural Philosophy Divided into Thirteen Parts: with an Appendix Containing five Parts/Written by the...Dvchess of Newcastle.* (Margaret Cavendish, 1668)

[30] […]; but of all men, Travellers have most reason to Adore and Worship God Best, for they see Most of his Wonderfull works, *which* shew his Power, Might, Wisdome, and Majesty, *the which* makes his Creatures Admire him, […] (Cavendish, 1662. EMMA)

[31] […]; but my Desire and Will was to make my Subjects Happy, to *which* end I bent all my Industry, *the which* I wish, my Successor may do the like, for good Subjects deserve a good Soveraign; […] (Cavendish, 1662. EMMA)

In both sentences, *the which* appears alongside plain *which*. I observed similar instances in MCEModESP (see section 4.1.2), where *which* referred to an undefined antecedent and the combined relativiser was, subsequently, used to provide additional information. In examples [30] and [31], these roles are not as apparent. The sentence in [30] is rather ambiguous and it seems like the pronouns refer to different antecedents. Furthermore, in example [31] *which* is used as a determiner. However, despite also being rather ambiguous, the antecedent in this sentence can be argued to be the same for both pronouns. If the referent is *to make my Subjects Happy* for both, then the previously spotted pattern is once again observed. Furthermore, *which* precedes *the which* in both cases. Should their referents be different in example [30], this would hint at a stylistic choice to avoid repetition.

In the same text, a significant rise can be spotted in the use of *who* and its inflected forms. The fact that more animate and/or personal referents were present in this specific text may be connected to its genre. The text *Orations of Divers Sorts Accommodated to Divers Places Written by the Lady Marchioness of Newcastle* is, indeed, the only non–scientific text of hers that I included in the analysis. Given the context *the which* is normally expected in, all texts should, ideally, have been scientific and/or medical. However, given the special focus on its diachronic development, it was necessary to include writing that was spread evenly over the authors' writing careers. As Cavendish wrote for a shorter period of time than the other three authors, the temporal distance between her texts was shorter and it was necessary to also include this non–scientific work in the study. This step was needed to cover the mid–1650s, the years around 1660 and the later 1660s. The following Table 4.6 displays the distribution of *which* and *the which* in Cavendish's analysed writing:

	Cavendish Text 1 (1655)	**Cavendish Text 2 (1662)**		**Cavendish Text 3 (1668)**
	which	*which*	*the which*	*which*
Personal reference (without pronouns)	0	7	0	2
Personal pronouns	0	0	0	0
Deity	0	4	0	0
Nonpersonal reference (without pronouns)	86	73	2	91
Nonpersonal pronouns	10	12	0	6
Total	*96*	*96*	*2*	*99*
Restrictive clause	15	16	0	14
Non–restrictive clause	81	80	2	85
Total	*96*	*96*	*2*	*99*
Subject	96	96	1	98
Object	0	0	1	0
Prepositional phrase	0	0	0	1
Total	*96*	*96*	*2*	*99*
Plural antecedent	11	19	0	30
Singular antecedent	64	43	0	24
Sentential antecedent	21	34	2	45
Total	*96*	*96*	*2*	*99*
Determiner	0	0	0	0
Others	3	3	0	1
Interrogative pronoun	1	1	0	0

Table 4.6: Grammatical profiles of which *and* the which *in Margaret Cavendish's selected texts. Design taken and adapted from Raumolin–Brunberg (Table 3; 2000: 212)*

The chi–squared tests that I carried out did not indicate any statistical relevance in any of the distributions. The fact that *the which* is present in Cavendish's writing is, however, interesting

in itself. First of all, the texts were all composed after 1650, which indicates a later usage of the relativiser than observed in other datasets. The Helsinki Corpus, for instance, does not feature any instances in texts from the second half of the 17th century (Rissanen 1999: 297). Furthermore, as mentioned at the beginning of this subsection, Cavendish's language seems to have changed noticeably depending on her social circle. Her exile in France and in the Netherlands ended in 1660, when she returned to London. Unfortunately, more data would be needed to confirm whether her use of *the which* may have been influenced by her social surroundings. With more of her writing being analysed, as well as a study of her husband's use of *which* and *the which*, perhaps it would be possible to draw more conclusions. Considering the course her use of third person singular *–s*/*–th* took, it may be possible that she did not use *the which* before her period in exile. Her husband, on the other hand, may still have used it, as he belonged to the previous generation. The two isolated occurrences of *the which* appeared in her text from 1662, composed only two years after her return to London. This could indicate that she accommodated her use of relativisers to that of her husband. Once the couple had returned to England, she may have adapted her language to that of the fellow Londoners, choosing plain *which* instead. If this is the case, it would account for the absence of *the which* in her 1668–text (but not for its lack in the one from 1655).

Another interesting observation about her writing is that *which* appeared several times with human and divine antecedents. In example [32], it is used with a human referent and a nonhuman, sentential antecedent (underlined). In example [33], on the other hand, divine subjects are referred to using both *which* and *who*:

[32] HEre is a man, *which* is Accused for Stealing privately, and Robbing openly, against all Law and Right, the Goods of his Neighbours, for *which* we have brought him before your Honours, […] (Cavendish, 1662. EMMA)

[33] […], the Belief Proceeds from the Son of God, *who* did Take upon Him the Shape of Man, but then we may believe, that Angels are of the Shape of Doves, because the Holy Ghost, *which* is Co–equal and Co–eternal with the Son, did Take upon Him the Shape of that Bird. (Cavendish, 1662. EMMA)

The fact that antecedents referring to persons and the Deity are mostly present in the text from 1662 may, once again, be traced back to its non–scientific genre. However, it is interesting how

the Son of God is referred to using *who*, while *which* is used with *the Holy Ghost*. Perhaps a differentiation was made between human–like and purely divine antecedents.

The text from 1668 also contained one single sentence [34] that featured both *that* and *which* (marked in italics):

[34] Wherefore, nothing can be a perfect, and a just Judg, but something *that* is Individable, and Unalterable, *which* is the Infinite GOD, […] (Cavendish, 1668. EMMA)

Just like in MCEModESP and in Boyle's texts, *which* is clearly preferred in non–restrictive relative clauses. What is more striking, however, is the fact that the pronoun is used almost exclusively in subject position. A preference for this role had already been observed in the first corpus (and in Raumolin–Brunberg's (2000) study), but not in Boyle's works. *The which* appeared one time as the subject and once with a preposition; both times in non–restrictive clauses.

Moving on, Cavendish frequently used *which* with plural, singular and sentential antecedents. In her 1655–text, there was a preference for singular referents. One of the antecedents occurring with *the which* was plural, while the other was sentential. Apart from two interrogative pronouns, the cases labelled as "others" in the last section refer to free relative clauses, as well as adverbial and adjectival antecedents.

Margaret Cavendish is the only female author included in this study and, also, the only female writer in EMMA. Previous studies on language change have reached significant conclusions on differences between genders (see, e.g., Raumolin–Brunberg 2000; Labov 2001 or Nevalainen & Raumolin–Brunberg 2003; 2016). For instance, Labov argues that women adopt prestige forms at a higher rate than men in language change from above. In changes from below the level of social awareness, women use innovative forms more frequently than men (1990: 213–15; 2001: 274; 292; cited in Nevalainen & Raumolin–Brunberg 2016: 111). Furthermore, *the which* was the option favoured by the female writers in Nevalainen & Raumolin–Brunberg's (2016: 129) study on language change in Tudor and Stuart England. This preference, which stood in contrast with their male informants, was noticeable until the middle of the 16th century. Therefore, it would be an interesting focus for a future study to compare her use of certain features with that of male authors featured in EMMA. In this particular project, however, this was not a central concern.

4.2.3 Frequencies and distribution of the relativisers in Nathaniel Crouch's texts

The next author in my analysis is Nathaniel Crouch (1640–1725). As mentioned in section 3.1.5, his profile differs from those of the other writers. Generally, the focus in this study lay on the language of members of the London Society. Within this context, it is natural that the other authors, as well as most authors included in EMMA, came from a privileged social background. Given the higher and more formal registers that *the which* was expected in (see section 2.1.4), this fit the focus of this project extremely well. However, another interesting question is that of whether there might be differences between individuals from different ranks of society. It is a well–established fact in sociolinguistics that factors such as education, gender, age, etc., often play an important role in linguistic change (Raumolin–Brunberg & Nurmi 2011: 252). For instance, the data from Johansson's (2012) study on relativisers in trial proceedings showed differences in the animacy distinction. Personal *which* was more present in witnesses' speech than in that of lawyers. While the members of the legal profession must have had a high formal education, the witnesses came from diverse social and educational backgrounds. Considering other factors, such as regional and dialectal variation, a direct correlation with the educational backgrounds could, however, not be established with certainty (785).

Similarly, in their study on defining in Early Modern English medical texts, McConchie & Curzan (2011: 91) observed differences between university–trained physicians and surgeons who had mainly gained their knowledge from practical experience. The former were often more precise and innovative in their writing, being in advance of the contemporaneous lexicographers. Surgeons, on the other hand, often lagged in defining abstract conceptual matters.

Considering the nature of historical corpora it is, unfortunately, rarely possible to study these differences in detail. Most datasets of older Englishes feature texts composed by upper– and middle–class male informants. The general shortage of writing produced by the lower social strata and women is due to the, until the end of the Early Modern period, still widespread illiteracy among the population (Nevalainen & Raumolin–Brunberg 2016: 26–7; see also section 2.2.3). Especially with a corpus like EMMA, where the focus lies on the elite ranks of society, it is not possible to form conclusions about the general public. It is also true, however, that in the context of academic and/or scientific texts a certain level of education is expected from the authors in any case.

Nathaniel Crouch grew up in a middle–class family of artisans. He did not attend university, but nevertheless pursued a successful career as a publisher in London (Fonteyn & Petré 2022: 98). Gaining status as a celebrated bookseller and writer at a young age, he assured himself a position among the London elite (Mayer 1994: 392–3). In a densely populated city such as London, it was a natural norm that individuals from all social ranks would interact with each other, at least to some degree. Social mobility was easier here than elsewhere, resulting in an influx of people from the surrounding rural villages. Consequently, London was a meeting point for people from a variety of socioeconomic backgrounds, resulting in a highly heterogenous population that saw daily exchanges among persons with different social roles (Nevalainen & Raumolin–Brunberg 2016: 39). Still, the case of this particular author is noteworthy. As observed by Mayer (1994: 393):

> "Crouch's career as historian and bookseller, […], constituted a relatively rare and therefore highly significant meeting ground for high and low culture in England at the end of the seventeenth century."

Contrary to the other three authors, Crouch did not work and write in the scientific or medical sectors. Instead, the works he published were mostly of religious or didactic nature. In his educational texts, he treated diverse historical topics, including contemporary events, histories of the Stuart kings, and an account of the Civil War. Within the context of his own upbringing and his later career, he is often described as a linking figure between readers of different social backgrounds. His works were popular among readers who would, normally, not have access to historical books and those who, in comparison, regularly consumed prestigious literature (Mayer 1994: 391–2; 394–5).

Shifting the focus to the language in his writing, it was difficult to predict whether any differences would be visible compared to the other authors. On the one hand, given results from previous studies (see above), there were grounds to believe that his limited access to higher education would be reflected in his texts. A feature like *the which*, typical of more formal writing, might have been absent from his works. On the other hand, as a celebrated publisher who quickly gained social status, it may well be that he accommodated his language to the higher register the elite members among his audience were accustomed to. Finally, it should be kept in mind that while Boyle and Cavendish belonged to the second generation in EMMA, Crouch was part of generation three. He lived and wrote towards the end of the Early Modern English period and in the early 1700s. Consequently, standardisation processes and the

disappearance of *the which* were mostly complete during his time. The total frequencies of the relativisers in his texts are displayed in Table 4.7:

Relative pronoun	**Crouch Text 1 (1678)[20]**		**Crouch Text 2 (1686)[21]**		**Crouch Text 3 (1696)[22]**	
	N	%	N	%	N	%
which	177	~72.5	427	~56.7	440	~45.9
the which	0	0	0	0	0	0
who	41	~16.8	215	~28.6	391	~40.8
whom	10	~4.1	59	~7.8	94	~9.8
whose	16	~6.6	52	~6.9	33	~3.4

Table 4.7: Total occurrences of the relative pronouns which*,* the which*,* who*,* whom *and* whose *in Nathaniel Crouch's selected texts*

As expected, Crouch did not use the relativiser *the which* in any of the three analysed texts, but continuously preferred other options. The use of *who*, on the other hand, increased over the decades. The antecedents *which* appeared with in his texts are displayed in Table 4.8. Once again, only randomly generated samples of 100 were included in the analysis:

[20] *Miracles of Art and Nature, or, A Brief Description of the Several Varieties of Birds, Beasts, Fishes, Plants, and Fruits of other Countreys: together with several other Remarkable Things in the World by R.B., Gent.* (Nathaniel Crouch, 1678)

[21] *A View of the English Acquisitions in Guinea and the East Indies with an Account of the Religion, Government, Wars, Strange Customs, Beasts, Serpents, Monsters, and other Observables in those Countries: together with a Dcription of the Isle of St. Helena and the Bay of Sculdania where the English usually Refresh in their Voyages to the Indies: Intermixt with Pleasant Relations and Enlivened with Picture/by R.B.* (Nathaniel Crouch, 1686)

[22] *The History of the Kingdom of Scotland. Containing an Account of the Original of that Nation, and of the most Remarkable Transactions and Revolutions during the Reigns of seventy two Kings and Queens, to the seventh year of King William III. Comprehending Variety of State Intreagues, Strange Accidents, Prodigies, and other Memorable Occurrences.: With a List of the Present Nobility of that Kingdom./By R.B.* (Nathaniel Crouch, 1696)

	Crouch Text 1 (1678)	**Crouch Text 2 (1686)**	**Crouch Text 3 (1696)**
	which	*which*	*which*
Personal reference (without pronouns)	4	0	0
Personal pronouns	0	0	0
Deity	0	0	0
Nonpersonal reference (without pronouns)	89	96	98
Nonpersonal pronouns	4	4	2
Total	*97*	*100*	*100*
Restrictive clause	15	8	7
Non–restrictive clause	82	92	93
Total	*97*	*100*	*100*
Subject	33	28	1
Object	36	54	50
Prepositional phrase	28	18	49
Total	*97*	*100*	*100*
Plural antecedent	42	28	15
Singular antecedent	45	52	24
Sentential antecedent	10	20	61
Total	*97*	*100*	*100*
Determiner	2	0	0
Others	1	0	0
Interrogative pronoun	0	0	0

Table 4.8: Grammatical profile of which *in Nathaniel Crouch's selected texts. Design taken and adapted from Raumolin–Brunberg (Table 3; 2000: 212)*

Personal *which* only appears a total of four times in Crouch's analysed writing, in his text from 1678. Along with the disappearance of *the which*, a compliance with the animacy distinction had been expected. Given that the texts were composed during the last decades of the period, referring to human antecedents using (*the*) *which* might have been perceived as inadequate, at least in higher registers (see, e.g., Nevalainen & Raumolin–Brunberg 2002: 120 or Sato 2016: 208).

Moving on to the next section of Table 4.8, the occurrences indicate a clear preference for *which* in non–restrictive relative clauses. Once again, Dekeyser's (1984: 69) claim of the pronoun assuming a 'neutral' role and being common in both clause types was not confirmed.

An interesting shift can be observed in the next section. In the first text, Crouch used *which* both as the subject and object of relative clauses, as well as in prepositional phrases. All these functions are still covered in 1686, but the distribution is less even. Instead, there seems to be a preference for *which* in object position. In the final text, the relativiser is used equally as object and in prepositional phrases, but only once as the subject. The chi–squared test that I performed in R resulted in a *p*–value of 9.06e–10. This value does, therefore, suggest that the null hypothesis (i.e., the relativisers not being used according to any system) can be rejected with high certainty.

In the next section, *which* appeared equally with both singular and plural antecedents in the first text and, to a lesser extent, with sentential ones. It is especially interesting how the sentential referents increased in his work from 1696 and became the most frequent antecedent type. The chi–squared test I performed resulted in a *p*–value of 1.142e–14. This distribution does, therefore, seem statistically relevant.

Finally, Crouch used *which* twice as a determiner in his 1678–text and once in a free relative clause. Otherwise, it never appeared as an interrogative pronoun and there were no unclassifiable cases. It seems, thus, that *which* had a strictly relative function in these three texts.

4.2.4 Frequencies and distribution of the relativisers in William Salmon's texts

The final author in my analysis was William Salmon (1644–1713). Alongside Robert Boyle (see section 4.2.1), his professional background fit the aim of this study the best. In the EMMA metadata, he is referred to as "doctor", even though it is not known whether he actually pursued university medical education. Other sources, rather, describe him as a "self–proclaimed

apothecary physician and London professor of physick", who advocated for access to medical care and literature for all (Apelbaum 2021: 41).

He was a famous figure in his field and had a reputation of being a brilliant publicist. By 1690, he had secured himself a notable position in the market of proprietary medicine. On the other hand, Kléber Monod (2013: 107; 129) describes him as "not much of a philosopher", given the straightforwardness of his writing and the simple instructions. As a consequence, established doctors were suspicious of his remedies, repeatedly questioning his practices (see also Apelbaum 2021: 47).

Salmon was born last of the four authors in this study. With his professional career starting after 1650, his writing was expected to be mostly in line with the conventions suggested by the 17th–century grammarians. For example, he was not expected to refer to human antecedents using *which*, but instead with other relativisers such as *who* (*whom*). Alternatively, he may have used *that* or the *zero* construction in this function, but these elements were not included in my study. The combined pronoun *the which*, on the other hand, was hardly expected in his data, considering that his earliest selected text was composed in the 1670s – a decade when *the which* had already largely fallen out of use. The total occurrences of the relativisers in Salmon's selected texts are showcased in Table 4.9:

Relative pronoun	**Salmon Text 1 (1672)[23]**		**Salmon Text 2 (1683)[24]**		**Salmon Text 3 (1695)[25]**	
	N	%	N	%	N	%
which	907	~88.5	1,666	~91	233	~93.6

[23] *Polygraphice, or, The Arts of Drawing, Engraving, Etching, Limning, Painting, Washing, Varnishing, Gilding, Colouring, Dying, Beautifying, and Perfuming in four Books: Exemplifyed in the Drawing of Men, Women, Landskips, Countries and Figures of Various Forms, the Way of Engraving, Etching, and Limning, with all their Requisites and Ornaments, the Depicting of the most Eminent Pieces of Antiquities, the Paintings of the Antients, Washing of Maps, Globes or Pictures, the Dying of Cloth, Silk, Horns, Bones, Wood, Glass, Stones and Metals, the Varnishing, Colouring and Gilding thereof according to any Purpose or Intent, the Painting, Colouring and Beautifying of the Face, Skin and Hair, the whole Doctrine of Perfumes, never Published till now, together with the Original, Advancement and Perfection of the Art of Painting / by William Salmon ...* (William Salmon, 1672)

[24] *Doron Medicum, or, A Supplement to the New London Dispensatory in III Books: Containing a Supplement I. to the Materia Medica, II. to the Internal Compound Medicaments, III. to the External Compound Medicaments: Compleated with the Art of Compounding Medicines.../by William Salmon...* (William Salmon, 1683)

[25] *The Family Dictionary, or, Houshold [sic] Companion wherein are Alphabetically Laid down Exact Rules and Choice Physical Receipts for the Preservation of Health...Directions for Making Oils, Ointments, Salves,...Chymical Preparations, Physical–Wines, Ales and other Liquors and Descriptions of the Virtues of Herbs, Fruits, Flowers...and Parts of Living Creatures used in Medicinal Potions,...likewise Directions for Cookery,...also the Way of Making all Sorts of Perfumes...together with the Art of Making all Sorts of English Wines,...the Mystery of Pickling and Keeping all Sorts of Pickles...: to which is Added as an Appendix the Explanation of Physical Terms, Bills of Fare...: with the Art of Carving and many other Useful Matters/by J.H.* (William Salmon, 1695)

the which	11	~1.1	63	~3.4	2	~0.8
who	47	~4.6	31	~1.7	11	~4.4
whom	15	~1.5	11	~0.6	0	~0
whose	45	~4.4	61	~3.3	3	~1.2

Table 4.9: Total occurrences of the relative pronouns which*,* the which*,* who*,* whom *and* whose *in William Salmon's selected texts*

Surprisingly, *the which* is indeed present in all of Salmon's texts. His work from 1672 counts 11 instances – not a large number, but significant considering that the use of the relativiser already declined in the 16th century (Raumolin–Brunberg 2000: 209; see also section 2.1.4). What is even more striking is its rise in the text from 1683. Composed at the end of the Early Modern English period, it still contained a total of 63 instances. This impressive number indicates that the pronoun may still have been in use later than previously assumed. The third text, written in 1695, is the second closest work to 1700 in my study (the latest text was Crouch's 1696–text, see section 4.2.3). Composed twelve years later than his 1683–work, Salmon's use of *the which* declined significantly, and the pronoun only appeared twice.

Who, as well as its inflected forms, were relatively rare in Salmon's texts. However, as Table 4.10 (see below) shows, (*the*) *which* was not frequent with personal antecedents either.

Regarding the genres of the texts, the works from 1683 and 1695 are classified as "science (medicine)" in EMMA. The text from 1672, *Polygraphice*, is an art manual. In the case of this author, too, a text that was not strictly scientific had to be included in the analysis to evenly cover his career. However, *Polygraphice* is a work where art meets alchemy and medicine, as it also contains recipes and guides to exotic practices such as palm reading. What characterises this text is Salmon's preference for plain, direct instructions, without any in–depth explanations of the abstract underlying concepts (Kléber Monod 2013: 107).

Considering the variety of the covered topics, it may well be that the earliest text contains more descriptions of humans and, therefore, more occasions to use *who* and *whom*. It is also in this text that example [35], featuring both *who* and *which* with animate antecedents (underlined), occurs:

[35] Honour is depicted with two wings on its shoulders; *which* as Alciatus saith, was made in the form of a little child, […], holding hand in hand the God Cupid, *who* leads the child to the Goddess Vertue, *which* is depainted right over against it. (Salmon, 1672. EMMA)

In example [35], the referents are divine rather than human. *The child* is referred to using *it* and, similarly, *Honour* is portrayed with wings on *its* shoulder. More peculiar, perhaps, is the fact that the God Cupid is described using *who*, but *which* is used with the Goddess Vertue. I do not have an immediate explanation for this variation. *Which* with human or divine antecedents was hardly expected in the later texts, and it is curious that Salmon would use it as a free variant of *who* with antecedents of the same category. Taking a closer look at the antecedents of *which* (a randomly generated sample of 100 was used here as well) and *the which*, their distribution is displayed in Table 4.10:

	Salmon Text 1 (1672)		**Salmon Text 2 (1683)**		**Salmon Text 3 (1695)**	
	which	*the which*	*which*	*the which*	*which*	*the which*
Personal reference (without pronouns)	0	0	0	0	0	0
Personal pronouns	1	0	0	0	0	0
Deity	0	0	0	0	0	0
Nonpersonal reference (without pronouns)	86	10	89	53	86	1
Nonpersonal pronouns	13	0	10	2	13	0
Total	*100*	*10*	*99*	*55*	*99*	*1*
Restrictive clause	26	0	24	0	23	0
Non–restrictive clause	74	10	75	55	76	1
Total	*100*	*10*	*99*	*55*	*99*	*1*
Subject	38	4	58	18	58	0
Object	28	6	21	33	21	1
Prepositional phrase	34	0	20	4	20	0
Total	*100*	*10*	*99*	*55*	*99*	*1*
Plural antecedent	13	5	41	4	8	0
Singular antecedent	67	5	40	41	69	1

Sentential antecedent	20	0	18	10	22	0
Total	*100*	*10*	*99*	*55*	*99*	*1*
Determiner	0	1	0	8	0	1
Others	0	0	1	0	1	0
Interrogative pronoun	0	0	0	0	0	0

Table 4.10: Grammatical profiles of which *and* the which *in William Salmon's selected texts. Design taken and adapted from Raumolin–Brunberg (Table 3; 2000: 212)*

As expected, personal (*the*) *which* did not occur in the samples from Salmon's selected works, except for one instance in his 1672–text where *which* was used with a personal pronoun (see example [36], where the referent is underlined). Otherwise, this role seems to have mainly been attributed to other relativisers.

[36] For to him that already understands it, the labour would be useless and unprofitable; to <u>him</u> *which* is already delighted therein, it would be needless and superfluous; […] (Salmon, 1672. EMMA)

Moving on, *which* occurs in both restrictive and non–restrictive relative clauses, but there is a clear preference for non–restrictive ones. This was the case in the great majority of all analysed texts. Even more striking is the distribution of *the which*. The combined pronoun appeared exclusively in non–restrictive relative clauses in Salmon's texts. This is very much in line with its original context, as described by Fischer, who stated that *the which* was mainly used in non–restrictive clauses in Middle English (1992: 303–4). Examples for this are displayed below in sentences [37] and [38], where the pronoun (marked in italics) appears in continuative clauses. The *p*–value I obtained for the 1683–distribution was 0.0001822 and, thus, statistically relevant. In the other two texts, it was above the established threshold.

[37] Of Sand there is various sorts, the chief are Higate Sand, and Tripoll; *the which* to make fit for the work you must order thus. (Salmon, 1672. EMMA)

[38] […] decant the clear Tincture, and put it to the former; *the which* keep in a Glass close stopp'd. (Salmon, 1683. EMMA)

Less in line with previous conclusions is the next section. In the data from her study, Raumolin–Brunberg observed a trend of *the which* occurring in prepositional phrases in Late Middle English. Especially the subject role was, instead, increasingly attributed to plain *which* (2000: 221). While *which* does, indeed, appear more often as the subject of the relative clause in the second and third texts, it is also common as the object and with prepositions. In the text from 1672, the distribution is rather even between subject positions and prepositional phrases. *The which*, on the other hand, is not popular with prepositions in this dataset. Apart from four instances in the 1683–text (opposed to 18 subject positions and 33 object functions), there are no occurrences of the relativiser in this role. In the first text, I obtained a *p*–value of 0.04033, while in the second it was 6.95e–06. In the third text the *p*–value did, instead, not indicate statistical relevance.

Plain *which* was popular with plural, singular and sentential antecedents in Salmon's texts. In the first and in the third texts, however, there is a preference for singular referents. In the 1683–text, the relativiser is distributed evenly between plural and singular antecedents, occurring also to a lesser extent with sentential ones. Once again, the distribution of *the which* does not confirm previous conclusions. According to Fischer's (1992: 303–4) claims, the pronoun should have been preferred with sentential antecedents. In the first and third of the selected texts, however, *the which* does not occur at all with this referent type. It does appear in this construction 10 times in the second text, which yielded the most data, but there is a clear preference for singular antecedents. In the first text, I obtained a *p*–value of 0.006715 and in the second a *p*–value of 1.721e–05. Especially the second distribution does, therefore, look highly relevant. No statistical significance was found in the third text.

The last section of Table 4.10 presents another interesting use of *the which*: in all three texts, Salmon had used the pronoun as a determiner. In the second and third texts, *which* further appeared in free relative clauses in the two instances labelled as "others".

Besides three instances where *that* and *which* appeared in the same sentence, *which* and *the which* also occurred together twice in the text from 1683. As was the case with MCEModESP, I cannot exclude that additional instances were overlooked, as these constructions were not an initial focus of my data collection. The structures in examples [39] and [40] are, however, similar to those observed in the first corpus and in Cavendish's writing (see sections 4.1.2 and 4.2.2):

[39] Now this is that we seek for, for our use in Medicine, with *the which* thou mayest make a safe progress in many Diseases, *which* will not yield to vulgar Medicines. (Salmon, 1683. EMMA)

[40] This black, is the Bird *which* flyes without Wings in the Night, *the which* even the first Coelestial Dew, hath by a perpetual Coction, and ascension and descension transmuted into the blackness of the Head of a Crow, *the which* doth afterwards assume the Peacocks Tail, […] (Salmon, 1683. EMMA)

Compared to the previous examples, the order of *which* and *the which* is reversed in example [39]. It was the only occurrence in my data where the plain relativiser did not occur in first position. Nevertheless, *which* appears in a restrictive relative clause, defining the otherwise rather vague referent *many Diseases*. *The which*, on the other hand, is part of a non–restrictive clause and preceded by a preposition. The two pronouns do, however, not refer to the same antecedent. A better example for this is sentence [40]: plain *which* is used in a restrictive clause with the referent *Bird*, adding information necessary to identify it. In the following clause, the combined relativiser refers to the same antecedent, but merely provides additional information. *The which* is used a second time in the sentence, but with a different referent.

4.2.5 Summarising the results from EMMA

In the first part of my study, I had taken a quantitative approach. However, the analysis of the second dataset, EMMA, was mostly qualitative (the relativisers in each text were, still, quantified). Further, I considered more sociolinguistic aspects when describing the data. My final research question, in fact, concerned the way contemporary authors used the relativisers in their writing. Special attention was also paid to the potential influence their socioeconomic backgrounds might have had on their language.

The employed corpus allows for large–scale studies of individual authors' writing careers (Petré et al. 2019: 85; see also section 3.1.4). However, given the scope of this project, the amount of data had to be narrowed down to just three to four texts per author.

A further distinction from the first half of the study was the timespan covered: the analysed texts were composed between 1655 and 1696. Therefore, only the last quarter of the Early Modern English period is represented in the selected writing and no data is available from earlier decades. The texts were composed when *the which* had, supposedly, mostly disappeared

from language users' pronoun inventory. Even Margaret Cavendish (1623–1673), who was born first of the four authors, was unlikely to have encountered and used the pronoun often. At this point, the question arises whether this study would have been more fruitful by analysing data from the previous century. After all, the Early Modern period presents the most variation between 1500 and 1600, and ongoing standardisation processes had largely reached stability by 1700 (Nevalainen & Raumolin–Brunberg 1989: 82; Fennell 2001: 138). Further, Raumolin–Brunberg (2000) focused on the Late Middle English period in her study, when *the which* was still a common option. The later timeframe in this project was set by the texts featured in EMMA. In the end, the later dataset provided valuable insights into later uses of *the which* and occasions to verify past conclusions.

Given that other datasets, such as the Helsinki Corpus, do not feature any instances of *the which* after 1650 (Rissanen 1999: 297), the relativiser was not assumed to be frequent in this dataset. Especially in the texts closest to 1700, hardly any instances were expected. However, two out of the four authors still used *the which* in their works. While Margaret Cavendish only used it twice, in her text from 1662, William Salmon used it 11 times in his 1672–text and an impressive 63 times in his work from 1683. What is curious is the fast decline in the following decade, as he only used the pronoun twice in his 1695–text. Considering the conclusions from past studies, these numbers would hint at a later use of *the which* than previously assumed. Furthermore, instances where *which* and *the which* were used in the same sentence appeared in both his and Cavendish's texts. These structures were also present in MCEModESP, providing further evidence for a concrete pattern where the pronouns performed different specialised functions (see section 4.1.2).

Looking into personal (*the*) *which*, especially Robert Boyle and Salmon seemed to comply with the animacy distinction and to attribute this role mainly to *who* (or, presumably, *that*). Cavendish and Nathaniel Crouch still occasionally used (*the*) *which* with personal or divine antecedents, but the occurrences were rare. My third hypothesis that the data would show a visible development was, thus, not confirmed, as already the texts closest to 1650 barely featured any instances of personal (*the*) *which*.

The final statement in my hypothesis, namely that the authors would comply early with the conventions suggested by the 17th–century grammarians due to their social status, is highly speculative. Overall, it seems like the four authors were indeed progressive in their use of the investigated relativisers. An apparent exception is Salmon's frequent use of *the which*. Considering its sporadic appearances in the other authors' writing, as well as in other corpora,

Salmon seemed to frequently employ an outdated and conservative option. Interestingly, it was his 1683–text that featured the most occurrences of the pronoun, and not his 1672–work. The readoption of an older pattern would coincide with Sankoff's (2018; 2019; see section 2.3.1) third trajectory type. More data would be needed to assess whether this really is a case of retrograde lifespan change.

Generally, it is difficult to determine the role of the authors' status and education. An overall compliance with standardisation processes and similar conventions was observed. However, as no writers of lower social backgrounds were included in my study, I am not able to draw any conclusions on this matter. Comparability is a frequent issue in historical corpus linguistics, given the lack of data produced by the lower ranks of society (Nevalainen & Raumolin–Brunberg 2016: 26–7). A more decisive factor than the writers' individual backgrounds may, ultimately, be the text genre itself. Early Modern English scientific writing was rather conservative and presented a high register (Taavitsainen 2000: 147). It can, therefore, not be excluded that features such as *the which*, which were no longer common in other genres, may still have been adequate at a later stage.

4.3 Reaching statistical significance

In the previous sections, I discussed the results from my study. The observed frequencies of the relativisers were displayed in tables, and I also illustrated the distribution of *which* and *the which* in terms of their antecedents.

Further, I calculated the *p*–value of each section with the help of R, indicating when it was below the established threshold of 0.05. In these distributions, it could be assumed that the rows and columns in the cross tables were not independent of each other. Within the context of this study, this means that the relativisers were most likely governed by a system. Their distribution was, thus, not by random chance. In other cases, however, this could not be safely stated. A potential reason for insignificant *p*–values may have been the, overall, rather small datasets that had been used. It cannot be excluded that analysing more data would have, eventually, resulted in a lower value.

R offers a simple way to test this. Using its assoc–function, I generated an association graph for the distribution of *which* and *the which* in restrictive and non–restrictive clauses in William Salmon's 1672–text. In theory, this operation could be carried out with any of the distributions; this example merely serves illustrative purposes. I chose this section of Table 4.10

because it features both pronouns, but the instances of *the which* are few and the chi–squared test that I performed resulted in *p*–value > 0.05. Entering the numbers from the distribution into R, I obtained the following graph:

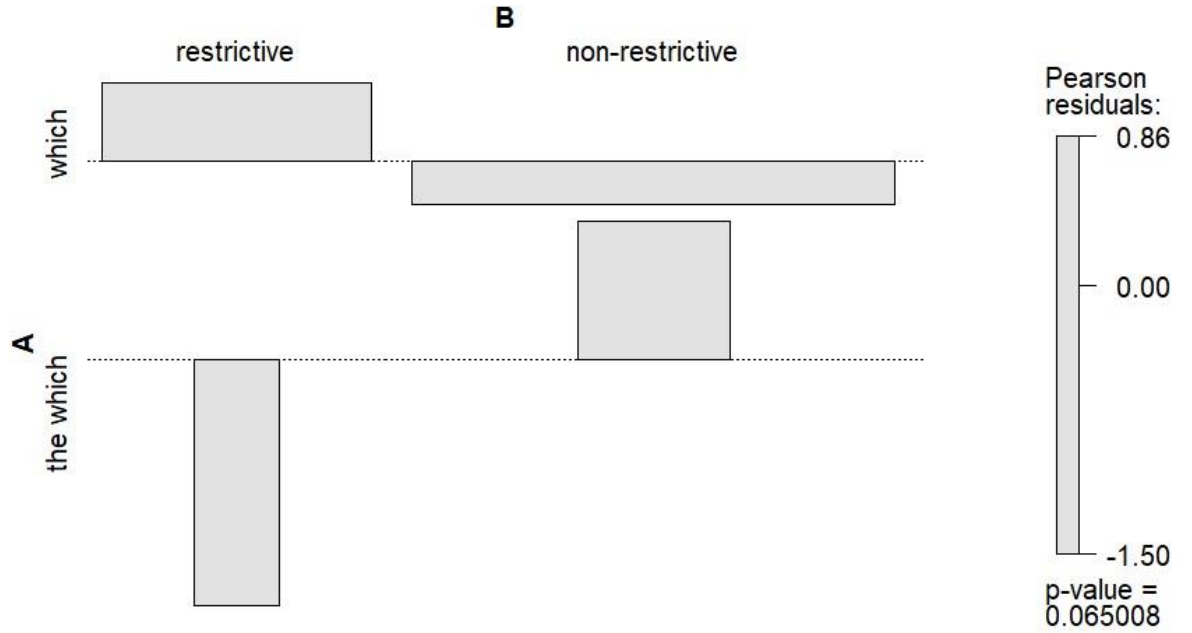

Figure 4.1: Association graph showing the distribution of which *and* the which *in restrictive and non–restrictive clauses in Salmon's 1672–text. The graph was generated using R's assoc function.*

A useful property of R's assoc function is its representation of each cell's Pearson residuals[26]. In Figure 4.1, this can be observed in the direction of the rectangles. Rectangles that grow above the baseline indicate that their observed frequency is greater than could be expected by chance (the residual is, thus, positive). If they grow below, their observed frequency is smaller than expected (the residual is negative). In other words, the direction of the rectangles indicates whether a variable is over or underrepresented in the data. Furthermore, the height of the bars shows their contribution to significance, while their width indicates their contribution to effect size[27] (Levshina 2015: 218–20; Johansson 2017: 28).

[26] In Levshina's words, the residuals in a contingency table indicate "the differences between the observed and expected frequencies divided by the squared root of the expected value" (2015: 218).

[27] The effect size is not related to statistical significance or *p*–values. Instead, it shows how strongly the different variables in a distribution are associated with each other, or to what degree groups of observations differ. A strong effect size does, therefore, not automatically mean that a distribution is statistically relevant (or vice versa) (Levshina 2015: 129; see also Johansson 2017: 24–5).

Figure 4.1 shows that restrictive *the which* is underrepresented in the data and has a low contribution to effect size (based on, respectively, the rectangle's height and width). This is not surprising, as there were indeed no occurrences in this text (see Table 4.10). Instead, the relative pronoun only appeared in non–restrictive clauses, resulting in an overrepresentation of this combination. Non–restrictive *which* had the highest number of occurrences (74 in total) and, consequently, also contributed the most to the effect size.

Figure 4.1 also shows that the *p*–value lies just above the threshold of 0.05. The situation would have looked different if a larger dataset had been analysed. By multiplying the numbers from the distribution by 10, I obtained the following graph:

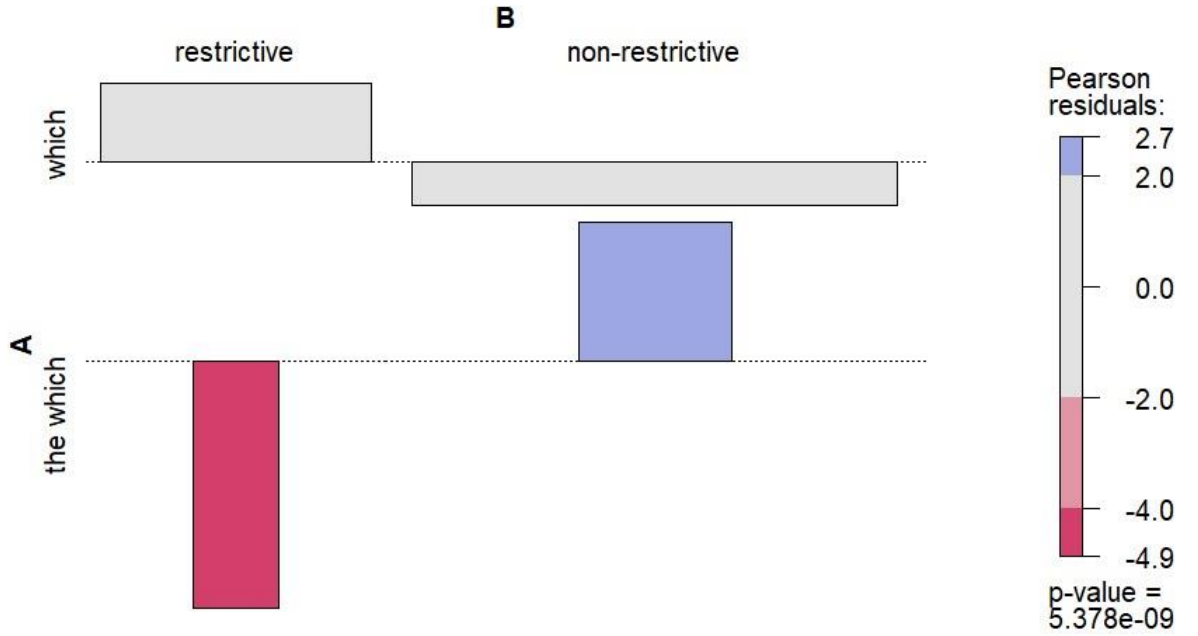

Figure 4.2: Association graph showing the distribution of which *and* the which *in restrictive and non–restrictive clauses in Salmon's 1672–text; the numbers were multiplied by 10. The graph was generated using R's assoc function.*

As the original *p*–value was already close to the threshold, statistical significance was already reached by doubling the data (*p*–value=0.0090673). This means that with twice as many observations, it would have been possible to safely reject the null hypothesis.

Compared to Figure 4.1, two of the rectangles have turned colourful in Figure 4.2. The colouring corresponds to a sign of a residual, and the intensity of its shade indicates its relative

importance (Levshina 2015: 219). Doubling the number of observations had been enough for restrictive *the which* to indicate a residual (and for its bar to, thus, turn colourful). Non–restrictive *the which*, on the other hand, turned colourful when the data was multiplied by 6. For illustrative purposes, however, I included a graph of the numbers multiplied by 10. The deeper shading and the lower *p*–value make it easier to describe the functions of the assoc graphing and how they can help to interpret linguistic data.

As mentioned above, twice as many observations would have been necessary to reach *p*–value < 0.05. However, more instances and a low *p*–value do not automatically equal relevance. In Johansson's words, "the results do not become […] better just because they are more significant". More observations may simply mean that the null hypothesis can be rejected with more confidence. But, in any case, it is very unlikely that the obtained frequencies were due to a truly random process (2017: 24–5).

The effect size, on the other hand, reveals more about the association between the variables themselves, or differences between groups of observations (see, e.g., Levshina 2015: 129 or Winter 2019: 159–61). A medium effect size[28], as Figure 4.2 shows in the case of restrictive *the which*, indicates practical significance. According to Johansson, therefore, it helps to evaluate how "useful" the obtained information is in practice (2017: 25).

Any deeper analysis would be beyond the scope of this project. For now, it is sufficient to estimate how much more data would have been needed to confirm the presence of patterns. In the case of this distribution from Salmon's 1672–text, doubling the observations would have been enough to reject the no association assumption. I employed this hypothetical test in R with other distributions as well. In most cases, far more instances would have been needed to define any underlying structures in the use of *the which*, as opposed to plain *which*.

As both inferential statistics and corpus linguistics deal with samples from real life populations (Levshina 2015: 7; Barth & Schnell 2021: 17), it is only natural that the available data is, generally, limited. Nevertheless, as Winter (2019: 159) points out, "bigger samples allow you to measure differences more accurately". For this reason, it is useful to hypothesise and test how large a dataset would need to be under ideal circumstances. This, in turn, helps to make generalisations about the results and may serve as a reference for future studies.

[28] According to Cohen's rule of thumb, effect sizes are considered small, medium, or large when they correspond to values of |0.2|, |0.5| and |0.8| or greater/less, respectively (1988). Measured in Pearson's r, the corresponding values range between -1 and +1 (the further away a coefficient is from zero, the stronger the correlation) (Winter 2019: 89; 160).

4.4 Summary

In this chapter, I presented and discussed the results from my study. In sections 4.1 to 4.1.2, I outlined the frequencies of the relativisers in MCEModESP. The resulting numbers, as well as a closer look at the distribution of *which* and *the which* in terms of their antecedents and clauses they occurred in, confirmed some previous conclusions and contradicted others. The most important finding, perhaps, were sentences containing both pronouns. Here, a pattern was observed that may suggest subtle differences in the use of *the which* as compared to plain *which*.

The same construction was observed in the data from EMMA, which was treated in sections 4.2 to 4.2.5. However, only two of the four authors still used *the which* in their writing. The general scarcity of the pronoun confirmed previous findings, as the texts stemmed from the second half of the 17th century, when its use had clearly declined. On the other hand, the observed occurrences suggest a later use of the relativiser than previously assumed.

Finally, in section 4.3 I illustrated how I further tested the distributions of the relativisers with the help of R's assoc function. With the help of graphs, I was able to estimate how much data would have been needed to reach p–value < 0.05. In the case of the selected example, twice as much data would have been sufficient. In other cases, however, far more evidence would have been needed to confirm any patterns.

5. Conclusion

The aim of this thesis was to investigate the development of relative *the which* during the Early Modern English period. Questions regarding the pronoun's origins, uses and disappearance during the 17th century remain largely unanswered by previous studies. For this reason, this project contributed to the discussion by testing past accounts and implementing suggested improvements.

As proposed by Raumolin–Brunberg (2000: 215), I took a step back from overarching conclusions and, instead, focused on one specific text genre. The genre, scientific prose, was mainly selected based on Fischer's (1992: 303–4) claims and was expected to yield an adequate number of occurrences of *the which* to perform this study. This approach proved to be fruitful, as especially MCEModESP provided numerous instances of the relativiser. Quantitative methods were used to analyse the data from this corpus.

Generally, previous grammatical accounts were largely confirmed. Exceptions were the antecedent type *the which* mostly occurred with (sentential referents were rather unpopular, as opposed to Fischer's (1992: 303–4) claim), and the fact that both the combined relativiser and plain *which* were preferred in non–restrictive clauses (according to Dekeyser (1984: 69), *which* was neutral in this regard).

A further improvement suggested by Raumolin–Brunberg (2000: 215) had been to employ multiple corpora, as past conclusions were often restricted to their specific dataset. For this reason, I added EMMA to my study. Working with a second corpus allowed me to compare observations made about MCEModESP, further testing their validity. Furthermore, the analysis of this corpus was mostly qualitative, and I was able to also include sociolinguistic aspects in my project.

Unfortunately, a comparison with authors of lower social ranks was not possible. Therefore, it is difficult to say whether the socioeconomic background of the writers in the corpus played a significant role in their use of the investigated relativisers. What stood out in the analysis, however, was the high individual variation. While some authors still included combined *the which* in their writing, others solely used plain *which*. Similarly, some authors still occasionally used (*the*) *which* with personal or divine antecedents, while others preferred *who* (or, presumably, *that*) in this role. Overall, each author presented highly interesting characteristics in their writing that are worth of further investigation.

Compared to MCEModESP, EMMA mostly contains texts from the second half of the 17th century. Furthermore, its metadata offers more precise dating of the texts. This allowed me to reach an important conclusion: combined *the which* seems to have been in use longer than previously assumed. Generally, the relativiser is believed to have disappeared around the middle of the Early Modern English period. Also, the Helsinki Corpus does not feature any instances after 1650 (Rissanen 1999: 297). Similarly, previous works rather focused on Late Middle English and only the first half the Early Modern English period when studying the pronoun (see, e.g., Raumolin–Brunberg 2000 or Nevalainen & Raumolin–Brunberg 2016). Analysing the selected texts from EMMA, however, especially William Salmon stood out. He still used *the which* over 60 times in his 1683–text and twice in his work from 1695. A possible reason for these later uses may lie in the rather conservative nature of medical writing (Taavitsainen 2000: 147). Also, as a doctor, Salmon may have used it for stylistic reasons to convey an additional sense of formality, professionalism, and elegance. These motivations are, naturally, highly speculative.

A second important observation regards the question of whether *which* and *the which* acted as free variants. In both MCEModESP and EMMA, I observed sentences (or sequences of sentences) which contained several different relativisers. Generally, higher registers in Early Modern English often featured long and complex sentences with multiple subordinate clauses (see, e.g., Lehto 2012; 2015; see also Görlach 1991: 124–25). Whenever a sentence contained both *that* and *which*, their performed functions were virtually identical. A simple explanation, thus, may be the authors' desire to avoid repetition. Further, it may have been a stylistic choice, meant to add variety to the texts and make them more attractive to the reader.

More interesting, however, were those instances where *which* and combined *the which* co–occurred (see section 4.1.2). Cases like these are not mentioned in any of the literature I consulted, but appeared in both corpora. In her study on the relativisers, Raumolin–Brunberg concluded that no real patterns were visible in her data. Consequently, she could not reject the hypothesis that *which* and *the which* acted as free variants, performing the same functions (2000: 221–2). My data, on the other hand, suggests that there may have been very fine differences in how the pronouns were used. When referring to the same antecedent, in fact, *which* always seemed to precede *the which*. In the examples found in the data, *which* occurred in a restrictive clause and introduced the referent. Once the antecedent had been defined, a non–restrictive clause would add additional, non–essential information. In these latter clauses, *the which* was used.

More observations would be needed to determine the role of *which* with certainty, and whether it only appeared in restrictive clauses in these constructions. In fact, I also observed instances where the same function was performed by *that*. The role combined *the which* assumes in this structure, however, seems fixed. Within the context of the information parameter, it may be that plain *which* (or *that*) is more neutral and suitable for vague, undefined antecedents. The combined relativiser, on the other hand, succeeds the previous clause(s) and builds a more concrete link to an identified referent. If the roles and the order were reversed, plain *which* or *that* might cause ambiguity (see, e.g., Fischer 1992: 303). Overall, this pattern seems too frequent and precise to be due to random chance, considering also that it appeared in both datasets in texts written by different authors.

As pointed out throughout the chapters, a clear limitation in this project was its overall scope. With the limited time at my disposal, it was only possible to analyse and discuss rather small portions of data. While in MCEModESP I was able to quantify the relativisers occurring throughout the corpus, in the case of EMMA I only included small fractions of the authors' writing in my study. Instead, the samples I worked with fulfilled a representative function. They revealed interesting insights into the development of the Early Modern English relative pronoun system. However, as I hypothesised and tested in section 4.3, more data would be needed to observe statistically relevant patterns. For this reason, an improvement for future studies may be to analyse larger datasets, ideally covering the entire Early Modern English period. For an even more exhaustive diachronic overview, Late Middle English writing could also be included in the study. This could prove especially interesting when studying sentences featuring both *which* and *the which*, as older texts may presumably contain more of these cases. Alternatively, especially a case study of Salmon's writing may reveal more insights into his late use of *the which*. Generally, a more in–depth investigation with EMMA might help to further define the timeline of the relativiser and its eventual disappearance.

6. References

Books and articles

Abbott, Edwin A. (1966[1870]). *A Shakespearean Grammar: An Attempt to Illustrate some of the Differences between Elizabethan and Modern English*. London: Macmillan [Reprint: New York: Dover (1966)]. Available at http://test.perseus.tufts.edu/hopper/text.jsp?doc=Perseus:text:1999.03.0080.

Andersen, Henning. (1973). *Abductive and Deductive Change*. In *Language*. Vol. 49(4), pages 765–93. Baltimore: Linguistic Society of America.

Anstey, Peter R. (2000). *The Philosophy of Robert Boyle*. Vol. 5. London: Routledge.

Anthonissen, Lynn. (2019). *Constructional Change across the Lifespan: The Nominative and Infinitive in Early Modern Writers.* In Bech, Kristin & Ruth Möhlig–Falke (eds.). *Grammar – Discourse – Context: Grammar and Usage in Language Variation and Change*. Pages 125–156. Berlin, Boston: De Gruyter. Available at https://doi.org/10.1515/9783110682564-005.

Anthonissen, Lynn & Peter Petré. (2019). *Grammaticalization and the Linguistic Individual: New Avenues in Lifespan Research*. In *Linguistics Vanguard: Multimodal Online Journal*. Vol. 5(2). Available at https://doi.org/10.1515/lingvan-2018-0037.

Anthony, Laurence. (2005). *AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom*. Presented at Professional Communication Conference 2005. In *IPCC 2005. Proceedings. International Professional Communication Conference*. Pages 729–737. Available at https://doi.org/10.1109/IPCC.2005.1494244.

Anttila, Raimo. (1989 [1972]). *Historical and Comparative Linguistics*. Current Issues in Linguistic Theory 6. 2nd Edition. Amsterdam: Benjamins [1st Edition New York: Macmillan].

Apelbaum, Natalie. (2021). Doron Medicum – *the Gift of Medicine: How a Rare 17th Century Medical Translation Positively Impacted the Community and Predicted Current Analgesic Techniques.* In *Anaesthesia and Intensive Care*. Vol. 49(1). Pages 41–50. Available at https://doi.org/10.1177/0310057X211038570.

Archer, Ian W. (2000). *Social Networks in Restoration London: The Evidence of Samuel Pepys's Diary*. In Shepard, Alexandra & Phil Withington (eds.). *Communities in Early*

Modern England: Networks, Place, Rhetoric. Pages 76–94. Manchester: Manchester University Press.

Ball, Catherine N. (1996). *A Diachronic Study of Relative Markers in Spoken and Written English*. In *Language Variation and Change*. Vol. 8(2), pages 227–258.

Barlow, Michael. (2013). *Individual Differences and Usage–Based Grammar*. In *International Journal of Corpus Linguistics*. Vol. 18(4), pages 443–478. Amsterdam: Benjamins. Available at https://doi.org/10.1075/ijcl.18.4.01bar.

Barth, Danielle & Stefan Schnell. (2021). *Understanding Corpus Linguistics*. London: Routledge.

Bately, Janet. (1965). Who *and* which *and the Grammarians of the 17th Century*. In *English Studies*. Vol. 46, pages 245–250.

Bell, Allan. (1984). *Language Style as Audience Design*. In *Language in Society*. Vol. 13(2), pages 145–204.

Biber, Douglas. (2009). *Corpus–Based and Corpus–driven Analyses of Language Variation and Use*. In Heine, Bernd & Heiko Narrog (eds.). *The Oxford Handbook of Linguistic Analysis*. Pages 159–192. Oxford: Oxford University Press. Available at https://doi.org/10.1093/oxfordhb/9780199544004.013.0008.

Birdsong, David. (2005). *Interpreting Age Effects in Second Language Acquisition.* In Kroll, Judith F. & Annette M. B. de Groot (eds.). *Handbook of Bilingualism: Psycholinguistic Approaches*. Pages 109–127. Oxford: Oxford University Press.

Birdsong, David & Jan Vanhove. (2016). *Age of Second–Language Acquisition: Critical Periods and Social Concerns*. In Nicoladis, Elena & Simona Montanari (eds.). *Bilingualism across the Lifespan: Factors Moderating Language Proficiency*. Pages 163–181. Washington, DC: American Psychological Association and De Gruyter Mouton.

Brinton, Laurel J. & Leslie K. Arnovick. (2006). *The English Language: A Linguistic History*. Oxford: Oxford University Press.

Brysbaert, Marc, Michaël Stevens, Paweł Mandera & Emmanuel Keuleers. (2016). *How many Words do We Know? Practical Estimates of Vocabulary Size Dependent on Word Definition, the Degree of Language Input and the Participant's Age*. In *Frontiers in Psychology*. Vol. 7, page 1116. Available at https://doi.org/10.3389/fpsyg.2016.01116.

Busse, Ulrich. (2012). *Pronouns*. In Bergs, Alexander & Laurel J. Brinton (eds.). *English Historical Linguistics: An International Handbook. Volume 1*. Vol. 34(1), pages 731–43. Berlin: De Gruyter Mouton.

Bybee, Joan L. (2006). *From Usage to Grammar: The Mind's Response to Repetition.* In *Language (Baltimore)*. Vol. 82(4), pages 711–33. Washington, DC: Linguistic Society of America. Available at https://doi.org/10.1353/lan.2006.0186.

Calle–Martín, Javier. (2021). *A Corpus–Based Study of Abbreviations in Early English Medical Writing*. In *Research in Corpus Linguistics*. Vol. 9(2), pages 114–130. Available at https://doi.org/10.32714/ricl.09.02.06.

Cheng, Winnie. (2012). *Exploring Corpus Linguistics: Language in Action*. Routledge.

Coates, Jennifer. (1993). *Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language*. 2nd ed. London: Longman.

Cohen, Jacob. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. Hillsdale: Erlbaum Press.

Cowie, Claire. (2012). *Morphology*. In Bergs, Alexander & Laurel J. Brinton (eds.). *English Historical Linguistics: An International Handbook. Volume 1*. Vol. 34(1), pages 604–20. Berlin: De Gruyter Mouton.

Cressy, David. (1980). *Literacy and the Social Order: Reading and Writing in Tudor and Stuart England*. Cambridge: Cambridge University Press.

Cressy, David. (2003). *Society and Culture in Early Modern England*. Aldershot: Ashgate.

Crombie, Alistair C. (1994). *Styles of Scientific Thinking in the European Tradition: The History of Argument and Explanation, Especially in the Mathematical and Biomedical Sciences and Arts*. 3 Volumes. London: Duckworth.

Crossgrove, William. (2000). *The Vernacularization of Science, Medicine, and Technology in Late Medieval Europe: Broadening Our Perspectives*. In *Early Science and Medicine*. Vol. 5(1), pages 47–63. Available at https://doi.org/10.1163/157338200X00047.

Curme, George O. (1912). *A History of the English Relative Construction*. In *Journal of English and Germanic Philology*. Vol. 11(1–3), pages 10–29; 180–204; 355–80. Illinois: University of Illinois Press.

Dekeyser, Xavier. (1984). *Relativizers in Early Modern English: A Dynamic Quantitative Study*. In Fisiak, Jacek (ed.). *Historical Syntax*. Pages 61–88. Berlin: De Gruyter Mouton. Available at https://doi.org/10.1515/9783110824032.61.

Diessel, Holger. (2019). *The Grammar Network: How Linguistic Structure is Shaped by Language Use*. Cambridge: Cambridge University Press.

Fennell, Barbara A. (2001). *A History of English: A Sociolinguistic Approach*. Vol. 17. Oxford: Blackwell Publishers.

Fischer, Olga. (1992). *Syntax*. In Blake, Norman (ed.). *The Cambridge History of the English Language: 1066–1476*. Vol. 2, pages 207–408. Cambridge: Cambridge University Press.

Fissell, Mary E. (2007). *The Marketplace of Print*. In Jenner, Mark S. R. & Patrick Wallis (eds.). *Medicine and the Market in England and Its Colonies, c. 1450–c. 1850*. Pages 108–32. Basingstoke and New York: Palgrave Macmillan.

Fonteyn, Lauren & Peter Petré. (2022). *On the Probability and Direction of Morphosyntactic Lifespan Change*. In *Language Variation and Change*. Vol. 34(1), pages 79–105. New York: Cambridge University Press. Available at https://doi.org/10.1017/S0954394522000011.

Franz, Wilhelm. (1924). *Shakespeare–Grammatik*. Heidelberg: Winter.

Franz, Wilhelm. (1939). *Die Sprache Shakespeares in Vers und Prosa: Unter Berücksichtigung des Amerikanischen Entwicklungsgeschichtlich Dargestell*. 4th Edition of *Shakespeare–Grammatik.* Halle: Niemeyer [Reprint Tubingen: Niemeyer (1987)].

Gries, Stefan Th. (2017). *Quantitative Corpus Linguistics with R: A Practical Introduction*. 2nd Edition. London: Routledge.

Görlach, Manfred. (1991). *Introduction to Early Modern English*. Cambridge: Cambridge University Press.

Hasselgård, Hilde, Per Lysvåg & Stig Johansson. (2012). *English Grammar: Theory and Use*. 2nd Edition. Oslo: Universitetsforlaget.

Hiltunen, Turo & Jukka Tyrkkö. (2011). *Verbs of Knowing: Discursive Practices in Early Modern Vernacular Medicine.* In Taavitsainen, Irma & Päivi Pahta (eds.). *Medical Writing in Early Modern English*. Studies in English Language. Pages 44–73. Cambridge: Cambridge University Press.

Hoffmann, Thomas. (2011). *Preposition Placement in English: A Usage–based Approach*. Cambridge: Cambridge University Press.

Hopper, Paul J. & Elizabeth Closs Traugott. (2003). *Grammaticalization*. 2nd Edition. Cambridge: Cambridge University Press.

Huddleston, Rodney & Geoffrey K. Pullum. (2005). *A Student's Introduction to English Grammar*. Cambridge: Cambridge University Press.

Johansson, Christer. (2017). *Elementary Statistics in the Cognitive Sciences*. Unpublished manuscript.

Johansson, Christine. (2012). *Early Modern English: Relativization*. In Bergs, Alexander & Laurel J. Brinton (eds.). *English Historical Linguistics: An International Handbook*. Vol. 34(1), pages 776–90. Berlin: De Gruyter Mouton.

Jones, Peter M. (2011). *Medical Literacies and Medical Culture in Early Modern England*. In Taavitsainen, Irma & Päivi Pahta (eds.). *Medical Writing in Early Modern English*. Studies in English Language. Pages 30–43. Cambridge: Cambridge University Press.

Kléber Monod, Paul. (2013). *Solomon's Secret Arts: The Occult in the Age of Enlightenment.* 1st Edition. New Haven: Yale University Press.

Kytö, Merja, Peter Grund & Terry Walker. (2007). R*egional Variation and the Language of English Witness Depositions 1560–1760: Constructing a "Linguistic" Edition in Electronic Form*. In Pahta, Päivi, Irma Taavitsainen, Terttu Nevalainen & Jukka Tyrkkö (eds.). *Towards Multimedia in Corpus Studies*. Vol. 2. Published online. Available at https://www.helsinki.fi/varieng/journal/volumes/02/kyto_et_at/.

Krulatz, Anna, Anne Dahl & Mona E. Flognfeldt. (2018). *Enacting Multilingualism: From Research to Teaching Practice in the English Classroom*. Oslo: Cappelen Damm Akademisk.

Labov, William. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. (2001). *Principles of Linguistic Change: Vol. 2: Social Factors*. Malden and Oxford: Wiley–Blackwell.

Lass, Roger. (1999). *Phonology and Morphology*. In Lass, Roger (ed.). *The Cambridge History of the English Language: Vol. III: Early Modern English 1476–1776.* Pages 23–155. Cambridge: Cambridge University Press.

Lehto, Anu. (2012). *Development of Subordination in Early Modern English Legal Discourse*. In Groom, Nicholas & Oliver Mason (eds.). *Proceedings of the Corpus Linguistics 2011 Conference, Birmingham, 20–22 July 2011*. Available at

http://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-176.pdf.

Lehto, Anu, (2015). *The Genre of Early Modern English Statutes: Complexity in Historical Legal Language*. Doctoral dissertation. Helsinki: University of Helsinki. Available at https://helda.helsinki.fi/items/70b5be8c-4274-440d-a865-9e5634400f89.

Leith, Dick. (1983). *A Social History of English*. London: Routledge & Kegan Paul.

Levshina, Natalia. (2015). *How to do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins Publishing Company.

Lightfoot, David. (1979). *Principles of Diachronic Syntax*. Cambridge Studies in Linguistics 23. Cambridge: Cambridge University Press.

MacIntosh, John J. (ed.). (2006). *Boyle on Atheism*. Toronto Studies in Philosophy Series. Toronto: Toronto University Press. Available at https://ebookcentral.proquest.com/lib/bergen-ebooks/detail.action?docID=4671250.

Mäkinen, Martti. (2011). *Efficacy Phrases in Early Modern English Medical Recipes*. In Taavitsainen, Irma & Päivi Pahta (eds.). *Medical Writing in Early Modern English.* Studies in English Language. Pages 158–79. Cambridge: Cambridge University Press.

Mayer, Robert. (1994). *Nathaniel Crouch, Bookseller and Historian: Popular Historiography and Cultural Power in Late Seventeenth–Century England*. In *Eighteenth–Century Studies*. Vol. 27(3), pages 391–419. Available at https://doi.org/10.2307/2739362.

McConchie, Rod & Anne Curzan. (2011). *Defining in Early Modern English Medical Texts*. In Taavitsainen, Irma & Päivi Pahta (eds.). *Medical Writing in Early Modern English*. Studies in English Language. Pages 74–93. Cambridge: Cambridge University Press.

McEnery, Tony & Andrew Wilson. (2001). *Corpus Linguistics: An Introduction*. 2nd edition. Edinburgh University Press.

McEnery, Tony & Andrew Hardie. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Meurman–Solin, Anneli. (2012). *Dialects*. In Bergs, Alexander & Laurel J. Brinton (eds.). *English Historical Linguistics: An International Handbook: Volume 1.* Vol. 34(1), pages 668–85. Berlin: De Gruyter Mouton.

Meyer, David, Achim Zeileis & Kurt Hornik. (2006). *The Strucplot Framework: Visualizing Multi–Way Contingency Tables with vcd*. In *Journal of Statistical Software*. Vol. 17(3), pages 1–48. Available at https://doi.org/10.18637/jss.v017.i03.

Mustanoja, Tauno F. (1960). *A Middle English Syntax*. Helsinki: Société Néophilologique.

Nevalainen, Terttu. (2006). *An Introduction to Early Modern English*. Edinburgh: Edinburgh University Press.

Nevalainen, Terttu. (2015). *Social Networks and Language Change in Tudor and Stuart London – only Connect?* In *English Language and Linguistics*. Vol. 19(2), pages 269–92.

Nevalainen, Terttu & Helena Raumolin–Brunberg. (1989). *A Corpus of Early Modern Standard English in a Socio–Historical Perspective.* In *Neuphilologische Mitteilungen*. Vol. 90(1), pages 67–111. Modern Language Society.

Nevalainen, Terttu & Helena Raumolin–Brunberg. (1994). *Its Strength and the Beauty of It: The Standardization of the Third Person Neuter Possessive in Early Modern English*. In Stein, Dieter & Ingrid Tieken–Boon Van Ostade (eds). *Towards a Standard English: 1600 1800*. Pages 171–216. Berlin: Mouton de Gruyter.

Nevalainen, Terttu & Helena Raumolin–Brunberg. (2002). *The Rise of the Relative* who *in Early Modern English*. In Patricia Poussa (ed.). *Relativisation on the North Sea Littoral*. *LINCOM Studies in Language Typology.* Vol. 7, pages 109–21. Munich: Lincom Europa.

Nevalainen, Terttu & Helena Raumolin–Brunberg. (2016 [2003]). *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. 2nd Edition. Routledge [1st Edition London: Longman].

Nevalainen, Terttu, Helena Raumolin–Brunberg & Heikki Mannila. (2011). *The Diffusion of Language Change in Real Time: Progressive and Conservative Individuals and the Time Depth of Change*. In *Language Variation and Change*. Vol. 23(1). Pages 1–43. Available at https://doi.org/10.1017/S0954394510000207.

Pahta, Päivi & Irma Taavitsainen. (2004). *Vernacularisation of Scientific and Medical Writing in its Sociohistorical Context.* In Irma Taavitsainen & Päivi Pahta (eds). *Medical and Scientific Writing in Late Medieval English*. Pages 1–22. Cambridge: Cambridge University Press.

Pahta, Päivi & Irma Taavitsainen. (2011). *An Interdisciplinary Approach to Medical Writing in Early Modern English*. In Taavitsainen, Irma & Päivi Pahta (eds.). *Medical Writing in Early Modern English*. Studies in English Language. Pages 1–8. Cambridge: Cambridge University Press.

Payne, John & Rodney Huddleston. (2002). *Nouns and Noun Phrases*. In Huddleston, Rodney & Geoffrey K. Pullum (eds.). *The Cambridge Grammar of the English Language*. Pages 323–524. Cambridge: Cambridge University Press.

Petré, Peter, Lynn Anthonissen, Sara Budts, Enrique Manjavacas, William Standing & Odile A.O. Strik. (2019). *Early–Modern Multiloquent Authors (EMMA): Designing a Large–Scale Corpus of Individuals' Languages*. In *ICAME Journal*. Vol. 43(1), pages 83–122. Walter de Gruyter.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Ratia, Maura & Carla Suhr. (2011). *Medical Pamphlets: Controversy and Advertising*. In Taavitsainen, Irma & Päivi Pahta (eds.). *Medical Writing in Early Modern English.* Studies in English Language. Pages 180–203. Cambridge: Cambridge University Press.

Raumolin–Brunberg, Helena. (2000). Which *and* the which *in Late Middle English: Free Variants?*. In Taavitsainen, Irma, Terttu Nevalainen, Päivi Pahta & Matti Rissanen (eds.). *Placing Middle English in Context*. Pages 209–25. Berlin: Mouton de Gruyter.

Raumolin–Brunberg, Helena. (2005). *Language Change in Adulthood: Historical Letters as Evidence*. In Nevala, Minna & Minna Palander–Collin (eds.). *European Journal of English Studies*. Thematic issue on Letters and Letter Writing. Vol. 9, pages 37–51.

Raumolin–Brunberg, Helena. (2009). *Lifespan Changes in the Language of three Early Modern Gentlemen*. In Nurmi, Arja, Minna Nevala & Minna Palander–Collin (eds.) *The Language of Daily Life in England (1450–1800)*. Pragmatics and Beyond New Series 183. Pages 165–96. Amsterdam: Benjamins.

Raumolin–Brunberg, Helena & Arja Nurmi. (2011). *Grammaticalization and Language Change in the Individual*. In Heine, Bernd & Heiko Narrog (eds.). *The Oxford Handbook of Grammaticalization*. Pages 251–62. Oxford: Oxford University Press.

Reuter, Ole. (1937). *Some Notes on the Origin of the Relative Combination* the which. In *Neuphilologische Mitteilung*. Vol. 38, pages 146–88.

Rissanen, Matti. (1999). *Syntax*. In Lass, Roger (ed.). *The Cambridge History of the English Language: Vol. III: Early Modern English 1476–1776*. Pages 187–331. Cambridge: Cambridge University Press.

Romaine, Suzanne. (1982). *Socio–Historical Linguistics: Its Status and Methodology*. Cambridge: Cambridge University Press.

Romero–Barranco, Jesús. (2020). *Linguistic Complexity across two Early Modern English Scientific Text Types*. In *Atlantis*. Vol. 42(2), pages 50–71. AEDEAN: Asociación Española de Estudios Anglo–Americanos. Available at https://doi.org/10.28914/Atlantis-2020-42.2.03.

Rowland, Caroline. (2014). *Understanding Child Language Acquisition*. London: Routledge.

Rydén, Mats. (1966). *Relative Constructions in Early Sixteenth Century English: with Special Reference to Sir Thomas Elyot*. Studia Anglistica Upsaliensia. Vol. 3. Uppsala: Almqvist and Wiksell.

Rydén, Mats. (1983). *The Emergence of* who *as a Relativizer*. In *Studia Linguistica.* Vol. 37, pages 126–34.

Saito, Toshio. (1961). *The Development of Relative Pronouns in Modern Colloquial English. A Statistical Survey of the Development of their Usage Seen in British Prose Plays from the 16th Century to the Present Time*. In *The Scientific Reports of Mukogawa Women's University.* Vol. (8), pages 67–89.

Sankoff, Gillian. (2018). *Language Change across the Lifespan*. In *Annual Review of Linguistics*. Vol. 4(1), pages 297–316. Annual Reviews.

Sankoff, Gillian. (2019). *Language Change across the Lifespan: Three Trajectory Types*. In *Language*. Vol. 95(2), pages 197–229. Washington: Linguistic Society of America.

Sato, Kiriko. (2016). *The Personal Use of Relative* which *in Shakespearean English: The Relevance of Social and Emotional Factors*. In *Anglia: Journal of English Philology*. Vol. 134(2), pages 207–38. Available at https://doi.org/10.1515/ang-2016-0024.

Sato, Kiriko. (2019). *The Relative* which *with Personal Antecedents in Shakespeare's History Plays*. In *Neophilologus*. Vol. 103(2), pages 273–91. Available at https://doi.org/10.1007/s11061-018-09591-5.

Schnell, Hildegard. (2009). *The Evolution of the English Scientific Register*. Munich: Grin Verlag.

Slack, Paul. (1979). *Mirrors of Health and Treasures of Poor Men: The Uses of the Vernacular Medical Literature of Tudor England*. In Webster, Charles (ed.). *Health, Medicine and Mortality in the Sixteenth Century*. Pages 237–73. Cambridge: Cambridge University Press.

Stein, Dieter. (1993). *Sorting out the Variants: Standardization and Social Factors in the English Language 1600–1800*. In Stein, Dieter & Ingrid Tieken–Boon van Ostade (eds.). *Towards a Standard English*. Pages 1–18. Germany: De Gruyter.

Steinki, Johannes. (1932). *Die Entwicklung der englischen Relativpronomina in Spätmittelenglischer und Frühneuenglischer Zeit*. Doctoral dissertation. Breslau: Die Schlesische Friedrich–Wilhelms–Universität.

Stirling, Lesley & Rodney Huddleston. (2002). *Deixis and Anaphora*. In Huddleston, Rodney & Geoffrey K. Pullum (eds.). *The Cambridge Grammar of the English Language.* Pages 1449–1564. Cambridge University Press. Available at https://doi.org/10.1017/9781316423530.018.

Taavitsainen, Irma. (2000). *Scientific Language and Spelling Standardization 1375–1550*. In Wright, Laura (ed.). *The Development of Standard English: Theories, Descriptions, Conflicts*. Pages 131–54. Cambridge: Cambridge University Press.

Taavitsainen, Irma. (2010). *Discourse and Genre Dynamics in Early Modern English Medical Writing*. In Taavitsainen, Irma & Päivi Pahta (eds.). *Early Modern English Medical Texts: Corpus Description and Studies*. Pages 29–54. Amsterdam: John Benjamins.

Taavitsainen, Irma & Päivi Pahta. (1995). *Scientific "Thought–Styles" in Discourse Structure: Changing Patterns in a Historical Perspective*. In Wårvik, Brita, Sanna–Kaisa Tanskanen & Risto Hiltunen (eds.). *Organization in Discourse: Proceedings from the Turku Conference*. Pages 519–29. Turku: University of Turku.

Taavitsainen, Irma, Peter M. Jones, Päivi Pahta, Turo Hiltunen, Ville Marttila, Maura Ratia, Carla Suhr & Jukka Tyrkkö. (2011). *Medical Texts in 1500–1700 and the Corpus of Early Modern English Medical Texts*. In Taavitsainen, Irma & Päivi Pahta (eds.). *Medical Writing in Early Modern English*. Studies in English Language. Pages 9–29. Cambridge: Cambridge University Press.

Taavitsainen, Irma & Jukka Tyrkkö. (2010). *The Field of Medical Writing with Fuzzy Edges*. In Taavitsainen, Irma & Päivi Pahta (eds.). *Early Modern English Medical Texts*. Pages 57–61. Amsterdam: John Benjamins.

van Gelderen, Elly. (2006). *A History of the English Language*. Amsterdam: John Benjamins.

Voigts, Linda. (1982). *Editing Middle English Medical Texts: Needs and Issues.* In Trevor Levere (ed.). *Editing Texts in the History of Science and Medicine.* Pages 39–68. New York: Garland.

Voigts, Linda. (1984). *Medical Prose*. In Anthony Edwards (ed.). *Middle English Prose: A Critical Guide to Major Authors and Genres*. Pages 315–36. New Jersey: Rutgers University Press.

Winter, Bodo. (2019). *Statistics for Linguists: An Introduction Using R*. New York: Routledge.

Zeileis, Achim, David Meyer & Kurt Hornik. (2007). *Residual–based Shadings for Visualizing (Conditional) Independence*. In *Journal of Computational and Graphical Statistics*. Vol. 16(3), pages 507–25. Available at https://doi.org/10.1198/106186007X237856.

Corpora and software

A Corpus of English Dialogues 1560–1760 (CED). (2006). Compiled under the supervision of Kytö, Merja (Uppsala University) & Jonathan Culpeper (Lancaster University).

Anthony, Laurence. (2023). *AntConc* (Version 4.2.2). Tokyo: Waseda University. Available at https://www.laurenceanthony.net/software.

Corpus of Early English Correspondence Extension (CEECE). Compiled by Nevalainen, Terttu, Helena Raumolin–Brunberg, Samuli Kaislaniemi, Mikko Laitinen, Minna Nevala, Arja Nurmi, Minna Palander–Collin, Tanja Säily & Anni Sairio. The Department of Modern Languages, University of Helsinki.

Early Modern English Medical Texts (EMEMT). (2010). Compiled by Taavitsainen, Irma, Päivi Pahta, Turo Hiltunen, Martti Mäkinen, Ville Marttila, Maura Ratia, Carla Suhr & Jukka Tyrkkö. Released on CD–ROM. Amsterdam: John Benjamins.

Early Modern Multiloquent Authors (EMMA) Corpus. (2015–2018). Compiled by Petré, Peter, Odile A. O. Strik, Lynn Anthonissen, Sara Budts, Enrique Manjavacas, William Standing & Emma–Louise Silva. Antwerp: University of Antwerp. Available at https://www.uantwerpen.be/en/projects/mind-bending-grammars/emma-corpus/

The Child Language Data Exchange System (CHILDES) Corpus. (1984–). Compiled by MacWhinney, Brian & Catherine Snow. Available at https://childes.talkbank.org/access/.

The Corpus of Early English Correspondence (CEEC). (1998). Compiled by Nevalainen, Terttu, Helena Raumolin–Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi &

Minna Palander–Collin. The Department of Modern Languages, University of Helsinki.

The Helsinki Corpus of English Texts. (1991). Compiled by Rissanen, Matti (Project leader), Merja Kytö (Project secretary), Leena Kahlas–Tarkka, Matti Kilpiö (Old English), Saara Nevanlinna, Irma Taavitsainen (Middle English), Terttu Nevalainen & Helena Raumolin–Brunberg (Early Modern English). The Department of Modern Languages, University of Helsinki.

The Málaga Corpus of Early Modern English Scientific Prose (MCEModESP) (2016–2020). Compiled by Calle–Martín, Javier, David Moreno–Olalla, Laura Esteban–Segura, Jesús Romero–Barranco, Teresa Marqués–Aguado, Jacob Thaisen & Hanna Rutkowska. Málaga: University of Málaga. Available at http://modernmss.uma.es.

Meyer, David, Achim Zeileis, Kurt Hornik & Michael Friendly. (2023). *vcd: Visualizing Categorical Data* (Version 1.4–12). R Package. Available at https://CRAN.R-project.org/package=vcd.

Parsed Corpus of Early English Correspondence. Parsed version. (2006). Annotated by Taylor, Ann, Arja Nurmi, Anthony Warner, Susan Pintzuk & Terttu Nevalainen. Compiled by the CEEC Project Team. York: University of York and Helsinki: University of Helsinki. Distributed through the Oxford Text Archive.

R Core Team (2021). *R: A language and environment for statistical computing* (Version 4.3.2 (2023–10–31)). Vienna: R Foundation for Statistical Computing. Available at https://www.R-project.org/.

Websites and online dictionaries

Aarts, Bas, Sylvia Chalker & Edmund S. C. Weiner. (2014). *The Oxford Dictionary of English Grammar*. 2nd Edition Online. Oxford: Oxford University Press. Available at https://www.oxfordreference.com/display/10.1093/acref/9780199658237.001.0001/acref-9780199658237. Last accessed on February 18th, 2024.

Anthony, Laurence. *AntConc Homepage.* On *Laurence Anthony's Website.* Available at https://www.laurenceanthony.net/software/antconc/. Last accessed on February 20th, 2024.

Baldick, Chris. (2015). *The Oxford Dictionary of Literary Terms*. 4th Edition Online. Oxford: Oxford University Press. Available at

https://www.oxfordreference.com/display/10.1093/acref/9780198715443.001.0001/acref-9780198715443. Last accessed on April 26th, 2024.

Early Modern English Medical Texts (EMEMT). On *Varieng: Research Unit for Variation, Contacts and Change in English*. Last updated on March 22nd, 2011, by Kauhanen, Henri. Available at https://varieng.helsinki.fi/CoRD/corpora/CEEM/EMEMTindex.html. Last accessed on April 26th, 2024.

Miranda García, Antonio, Javier Calle–Martín, David Moreno Olalla, Santiago González Fernández Corugedo & Graham D. Caie. (2007–2016). *The Málaga Corpus of Early Modern English Scientific Prose*. Available at https://modernmss.uma.es/. Last accessed on February 19th, 2024.

Peters, Pam. (2013). *The Cambridge Dictionary of English Grammar.* Cambridge: Cambridge University Press. Available at https://www.cambridge.org/core/books/cambridge-dictionary-of-english-grammar/CE0A0355C932DAB20A4B48B88CC67BD9. Last accessed on May 9th, 2024.