Morphosyntactic variation and change in Late Modern English. A sociolinguistic perspective

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Philosophie des Fachbereiches 05
der Justus-Liebig-Universität Gießen

vorgelegt von
Bianca Widlitzki

aus Staufenberg

2018

Dekan: Prof. Dr. Thomas Möbius
1.Berichterstatter: Prof. Dr. Magnus Huber
2.Berichterstatter: Prof. Dr. Joybrato Mukherjee
Tag der Disputation: 20. Juni 2018

Table of contents

Erklärung zur Dissertation

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Bianca Widlitzki
Gießen, den 15. Januar 2018

Acknowledgements

The years in which I worked on my PhD project were shared with many people who have had an impact on both my professional and my personal life. I would like to take this opportunity to say thank you to some of them here.

First of all, I would like to thank Magnus Huber, who not only accompanied my research as my PhD advisor but also first opened the doors to the world of linguistics for me when he hired me as a student assistant for the APiCS project in 2010. Many thanks are also due to Danny Mukherjee, in whose team I've worked since 2013 and who agreed to be the second reviewer for my thesis.

I am very grateful to all my colleagues – researchers, technicians and administrative staff, past and present – at the English Linguistics section at Justus-Liebig University Giessen. As you are many, I hope you'll forgive me for not listing each of you individually by name. All of you taught me something – whether that was how to wrangle huge data tables with thousands of rows, what to watch out for when presenting or teaching, and many other professional skills… I profited immensely from the opportunity to watch and learn from you. More importantly, you were an inspiration beyond the purely work-related. I always appreciated about our team that there was a great deal of support for each other, especially when times were tough.

My short-term colleagues in Regensburg also deserve a shout-out: you made me feel very welcome in the couple of months I substituted for one of your own. A big thank you is also due to my former (APG) and present (admin) colleagues at the university outside the Department of English. You offered me a valuable outside perspective on my work, and your kindness and support towards the end of the dissertation writing process means a lot to me.

I am especially grateful to the other PhD students at the Giessen Department of English – it was invaluable to be able to discuss with you the everyday challenges, roadblocks and small triumphs of the research process. Special thanks go to my mentoring group in the SciMento programme (you are fantastic, ladies!), my MEWISMA mentor Nele, and of course my no. 1 dissertation writing buddy, Barbara Güldenring from the University of Marburg. Barbara invited me to write in her office once a week after we met at a conference in 2015, and we've been friends ever since. I am very grateful to Barbie and to my friend Julia, whom I've known since we were

both students in Giessen, for offering thoughtful commentary on the content and form of earlier versions of this thesis.

Most importantly, my friends and family have my deepest gratitude: your encouragement and support mean the world to me. Among this group (you know who you are), some people stand out: my brother Timm and my practically-sister-in-law Kathi (as far as I am concerned, my sister in any of the ways that count), you are the best! Thanks for having my back. Mama und Papa, euch möchte ich besonders herzlich danken. Als ich noch in der Schule war, sagte einmal eine Klassenkameradin zu mir: „Ach, Bianca, du könntest alles hinschmeißen und sagen, dass du freischaffende Künstlerin wirst, und deine Eltern würden trotzdem versuchen, es zu verstehen, und würden dich unterstützen.“ Auch wenn sich die Vorhersage zu meiner Berufswahl (zum Glück?) nicht bewahrheitet hat, stimmt der zweite Teil definitiv. DANKE.

List of figures

List of tables

1 Introduction

> [Mary Ann Newton:] I told him, as near as I could, an outline of the evidence that I wished to give.
> (OBC-Proc, t18630608-856)

1.1 Motivation and aim

Research in recent decades has gone a long way towards advancing linguistic analysis of the long-neglected Late Modern English period (roughly 1700-1900). Among other things, this is evidenced by several book-length overviews (Bailey 1996 and Görlach 2001 for the 18th, Görlach 1999 for the 19th century) and the publication of introductory textbooks for the period (Beal 2004, Tieken-Boon van Ostade 2009). The recurring *International Conference on Late Modern English*, established in 2001, as well as various edited volumes on diverse issues (recent examples include Dossena & Tieken-Boon van Ostade 2008 on correspondence, Hickey 2010b on ideology and change in 18th-century English, and Hundt 2014b on syntax) also attest to the recent heightened attention paid to the subject.

Despite these efforts, the picture of Late Modern English (abbreviated: LModE) remains surprisingly patchy, considering the comparatively large amount of textual evidence that exists for the period. Only some years ago, Anderwald (2012: 28) still called the 18th and 19th centuries the stepchildren of historical linguistics, and Aarts et al. (2012: 883) lamented a "descriptive gap between the Early Modern English and Present-day English period". Indeed, these criticisms are still valid, especially with regard to some areas of research.

Sociolinguistic research on Late Modern English is one such area. At a time when the field of historical sociolinguistics was only on the brink of emerging as a separate discipline, Rydén (1979: 34–35) wrote about a "pressing need for studies of socially conditioned syntactic variation in the last centuries". Much more recently, Smitterberg (2012: 953) called the sociolinguistics of Late Modern English "underresearched" still and stressed the necessity to do further work in this area. Two aspects in particular seem to require additional attention, namely studies involving

language from all social classes and studies providing systematic quantitative data on socially conditioned variation and change throughout the period.

That these highly relevant aspects still remain – at least to some extent – desiderata for Late Modern English is due to several factors. As for the representation of social classes, it is generally problematic for historical linguists that the written record of a period mostly reflects the language of high-status groups. For Late Modern English, access to literacy and opportunities to produce printed texts was heavily skewed in favour of higher-class men (see 2.2.1). As most corpora and other collections used for linguistic analysis are based on published material that exhibits this kind of bias, this poses a problem for detailed sociolinguistic analyses that aim to involve all strata of society.

The issue is compounded if more than one dimension (e.g. social class and gender) is considered. Difficulties arise even with relatively frequent phenomena: Smitterberg's (2005: 4) study on the progressive in Late Modern English, for instance, had to exclude the factor social class because the ca. 1-million word Corpus of Nineteenth Century English (CONCE, Kytö et al. 2000) did not contain enough material by lower-class speakers. Using larger corpora does not necessarily solve that problem: while they admittedly offer a greater chance of retrieving an adequate number of instances of a particular feature, their design is often not intended to support in-depth research on social variation. For example, the compilers of the Corpus of Late Modern English Texts (CLMET), which contains 10 million words, caution that their corpus is "unfit for any fine-grained sociolinguistic analysis" (de Smet 2005: 78). Many larger historical corpora like the CLMET or ARCHER (A Representative Corpus of Historical English Registers) are primarily meant to serve other purposes, such as studies of genre development. There is thus a need for studies involving all strata of Late Modern society, and especially such that incorporate the language of the lower classes (Kytö & Smitterberg 2006: 226)

With this in mind, it comes as no surprise that many of the significant insights on social variation in Late Modern English are indeed furnished by studies that highlight the language of individuals and social networks (frequently literary networks) and can make do with smaller samples of data, frequently in the form of letter collections. Research on the "Bluestocking Letters" (e.g. Sairio 2006, Sairio

2009) and on the correspondence of the "Southey-Coleridge Circle" (Pratt & Denison 2000) may serve as examples. In order to place these valuable contributions into a larger context, it is essential to pursue more quantitatively oriented approaches. Such large-scale systematic studies are rare, though:

> [t]here are few systematic historical investigations of language changes in their social contexts to date. This means that in many cases we have no knowledge of such basic issues as the time courses of the changes that took place in English over a given period of time. (Nevalainen & Raumolin-Brunberg 2003: 20–21)

Therefore, Nevalainen & Raumolin-Brunberg (2012: 24) call for studies providing "quantitative baseline data", i.e. data from a large reference group, to allow sociolinguists to connect more qualitatively oriented work on individuals' language choices with the overall development of features through time and thus connect the macro- and the micro-level of analysis.

The present study is motivated by the desire to provide one such large-scale systematic investigation of variation and change in Late Modern English from a sociolinguistic perspective. The focus on morphosyntax also aims at addressing existing research gaps:

> Despite the recent progress in the historiography of the English language between 1700 and 1900, morphological and syntactic change in LModE is still the least researched aspect of this period. (Hundt 2014a: 1)

The aim of this study is to add more information on these neglected aspects and, in doing so, to complement earlier research. A detailed outline, including the features under investigation, objectives and methodology is provided in 1.2.

1.2 Scope and methodology

This study examines morphosyntactic variation and change in Late Modern English from a sociolinguistic perspective. The objective is to shed light on the social dimension of the development of select linguistic features in the verb phrase throughout a period of almost 200 years. Additionally, it assesses the usefulness of trial proceedings in historical sociolinguistic research.

The study will make use of a broad empirical basis, employing the Old Bailey Corpus (OBC 1.0, Huber et al. 2012) as a primary source of data. The OBC consists of trial proceedings between 1720 and 1913. It avoids many of the problems of other Late Modern corpora as it was designed from the start with sociolinguistic applications in mind: it contains data from persons of all ranks and the relevant sociolinguistic annotation, and is quite sizeable with its 14 million words, which should allow for a fine-grained analysis of linguistic features. The court transcripts also provide a glimpse of the language of people who may not have been able to write themselves and leave written records. This source of data thus presents an excellent opportunity for a large-scale study of variation and change in Late Modern English. The Corpus of Late Modern English Texts 3.0 (CLMET 3.0, Diller et al. 2011) is used to provide additional data, conduct inter-genre comparisons and integrate a broader perspective.

Rather than covering a large amount of features in a cursory manner, the focus is an in-depth analysis of the four morphosyntactic features presented in Table 1.

linguistic domain	**feature**	**(major) variants**
modality	1) verbal expressions of obligation and necessity	HAVE TO: *You **have to** go.* MUST: *You **must** go.*
auxiliation	2) choice of perfect auxiliary	BE + past participle: *I **am come** home.* HAVE + past participle: *I **have come** home.*
tense	3) choice of narrative tense with SAY	historic present: *so **I says** to her* past tense: *so **I said** to her*
agreement	4) verb form with 2SG pronoun *you*	*was:* ***you was** alone* *were:* ***you were** alone*

Table 1. Morphosyntactic features investigated in the present work

The features cover different aspects in the verb phrase and have not yet been analysed in detail from a variationist-sociolinguistic point of view for the Late Modern period. Features 1 and 2 undergo change in the 18th and 19th centuries, while features 3 and 4 include one variant each that is primarily associated with spoken or conversational

style (the historic present form *I says* is a feature of spoken narrative; *you was* is considered a conversational feature at least in contemporary English).[1] This diversity allows for the investigation of different aspects of variation and change processes in LModE.

Each feature is analysed with regard to the following three questions:

A. How do variation and change in selected morphosyntactic features manifest themselves in Late Modern English with regard to the timing of change (if present) and its social and linguistic factors? How do different speech-related genres compare?
B. How are the variants evaluated in grammars of the time (positive / negative / changing)? Is there a correlation between this evaluation and use?
C. How suitable are the *Proceedings of the Old Bailey* (and trial proceedings in general) for historical sociolinguistics? How close to the dialogue uttered in the courtroom can we assume these published transcripts to be? What needs to be taken into account (e.g. in terms of scribal/editorial interference) when basing linguistic analyses on trial proceedings?

These questions are of course interrelated and difficult to discuss divorced from one another. Nevertheless, I will at this point outline what each of the three aspects entails.

Aspect A is perhaps the most straightforward. The interest here is on the diachronic development of selected linguistic features in Late Modern English, with a focus on the influence of social factors. The analysis entails questions like these: Which variables are socially salient and socially conditioned in the first place? Are the same population groups in the lead whenever change occurs? These questions are discussed based on evidence from the Old Bailey Corpus and under consideration of multiple independent variables that may have an impact on their distribution. To assess the situational impact of the courtroom setting, the overall development of the features in question in the OBC is compared to a subcorpus of the CLMET that is also speech-related but not operating under the same situational constraints.[2]

Aspects B and C address the broader setting in which the texts in the OBC were produced. Their inclusion is essential to ensure that the historical context is taken

[1] Further selection criteria that the features had to fulfil are presented in Chapter 3.4.
[2] The drama subcorpus of the CLMET is used as a reference corpus for three of the features; the narrative fiction subcorpus is used for one feature because there were not enough tokens for analysis in the drama corpus.

into account and thus for the success of this diachronic investigation. Aspect B is concerned with how the (variants of) linguistic features under investigation were evaluated by society at large; aspect C is focussed on the smaller sphere of the *Proceedings of the Old Bailey* and the circumstances of their production.

It is well known that the standardisation and codification of English was a major development in the period under investigation (see 2.3.2 and 2.3.3). In the wake of this development, social evaluation was attached to linguistic choices, which, in turn, led many speakers who wanted to advance in society to seek out instruction on how to use English 'correctly'. Guidance was provided in a plethora of pronunciation manuals, dictionaries and grammars. It is therefore necessary to assess in how far ideas of prestige and norms of correctness apply to the variants under investigation. This falls into the realm of question B, which takes the grammarians' point of view into account by drawing on information from 18th and 19th century grammars. It should be interesting to see whether tendencies observed in earlier work, i.e. that social climbers frequently use the prestigious emerging standard variants, can be confirmed for the features investigated here. Combining the study of grammars with the study of real-time developments is also intended to further our understanding of the relation between prescription and use: can a case be made that prescriptions affected language choices, or did grammarians primarily record what was already best practice?

Aspect C, finally, is concerned with the primary data, i.e. *The Proceedings of the Old Bailey*. Trial proceedings and similar so-called speech-based records are of course not to be equated with the spoken language of the time. Rather, they are passed down to us along with several 'filters' (Schneider 2013: 58, 60) that are created by the transfer to the written medium. Part of our task as historical linguists is to identify these filters and account for them in our interpretations. In the present case, we are dealing with a publication with a long history of over 230 years, created in a sophisticated process involving e.g. scribes and printers, and shaped by several interest groups and a changing legal system. It is therefore crucial to address the reliability of the *Proceedings* as a source for linguistic inquiry, and to identify historical events and actors that might influence the particular genre of trial proceedings.

On the whole, these three questions take into account social (A, B), situational (B) and textual (C) factors, as illustrated in Figure 1.

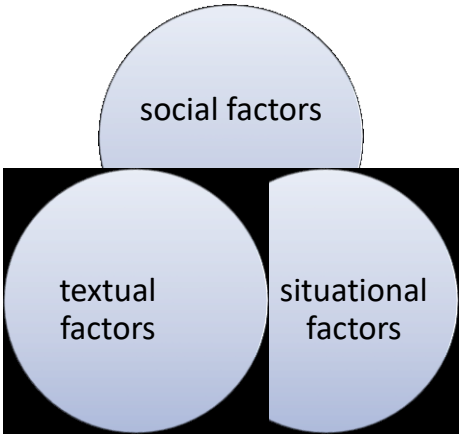


Figure 1. Areas of interest for the study: social, situational and textual factors in language variation and change

More often than not, they cannot be strictly separated from each other: for instance, a change in editing policy (which represents the textual dimension) may be triggered by changing norms of propriety, i.e. with regard to the detail of crimes recorded (located in the social dimension). These areas are therefore best conceptualised as overlapping and best discussed in an integrated manner in order to present a differentiated picture of English in the 18th and 19th centuries.

It is hoped that the study will play a part in a comprehensive analysis of Late Modern English and add to theoretical discussions of the impact of social factors in language change. In addition, the critical evaluation of Late Modern trial proceedings as a source for historical (socio)linguistics will be of interest to users and compilers of historical corpora. The language of court proceedings and depositions is frequently used in historical corpus linguistics (e.g. Moore 2008 or Kytö et al. 2011), often with the intention to gain access to a speech-related style. An assessment of Late Modern trial proceedings is thus called for: questions of authenticity and genre conventions as well as the potentials and limitations of this resource need to be addressed.

1.3 The upcoming chapters

The present work consists of three components: the frame, which is provided by this introduction and the summative presentation of results and concluding remarks in Chapter 8, a section on theory and methodology (Chapters 2-3) and the discussion of the corpus analysis (Chapters 4-7).

Chapters 2-3 address the theoretical background of this study and key methodological concerns. Chapter 2 outlines the theoretical foundations of historical sociolinguistics, discusses the role of social factors in language variation and change and locates the present study in the broader context of the field. General aspects about studying past language use are in the foreground, while the discussion of Late Modern linguistic developments is kept rather brief. As the variables selected for analysis are quite diverse, their specific histories and characteristics are relegated to individual chapters in the second (analytical) part of the present work. Importantly, the chapter includes information on the external history of the 18th and 19th centuries and surveys the changing attitudes to language in the era. Chapter 3 is concerned with the empirical basis of the study and the methodology. The Old Bailey Corpus and the Corpus of Late Modern English Texts are introduced. Especially the Old Bailey Corpus, which is the main source of data, is discussed in detail: its characteristics as a linguistic corpus and the history of the underlying publication are addressed. The linguistic and social variables selected for investigation are introduced and elaborated on before the chapter closes with a description of the analytical procedure.

The discussion of the corpus analysis is divided into four case studies (Chapters 5-8). The individual case studies on modality (Chapter 4), perfect auxiliary choice (Chapter 5), narrative tense (Chapter 6) and agreement (Chapter 7) are all largely structured in the same way. They each contain a description of the variable, a summary of previous research and of contemporary grammarians' comments on the feature, a review of methodological concerns and finally the discussion of the results of the corpus analysis. Chapter 8 serves as a conclusion, containing a summary of key results as well as an in-depth discussion of their theoretical and methodological implications: the research questions introduced in Chapter 1 are answered, concluding remarks are made and avenues for further research are pointed out.

1.4 Typographical and citation conventions

Language examples from the corpora are numbered; emphases in the language examples are mine unless otherwise stated. Emphases in quotations from other works, whether in the form of italic or boldface, are taken over from the original unless otherwise stated. Linguistic material is presented in italics, and small capitals stand for all forms of a verbal paradigm: for instance, HAVE encompasses *has*, *have*, *had*, and *having*. Capital letters are used for variables/factors in regression models (e.g. SOCIAL CLASS or MAIN VERB). Where percentages or probabilities are given, they are rounded to whole tenths. An exception to this rule holds for the regression models (for details, see 3.5), where estimates, standard errors, z-values and confidence intervals are not rounded according to this rule, but presented as they appeared in the model output.

Quotations from the corpora will be accompanied by either file name (for the CLMET) or trial identifier (for the OBC). Trial identifiers consist of two components, the date of the issue of the *Proceedings* in which the quote appears (in the format yyyymmdd) and the trial number. Trial numbers exist for all trial accounts in which spoken dialogue is featured. For instance, *OBC, t17790217-28* refers to trial no. 28 in the *Proceedings* of 17 February 1779. At times, I may quote from issues of the *Proceedings of the Old Bailey* that were not incorporated into the Old Bailey Corpus These examples will feature the abbreviation *OBCProc* instead of *OBC*. When quoting accounts of the *Proceedings* that do not include spoken dialogue and thus have no trial number, the date of publication (in the format yyyymmdd) will indicate the provenance of the quoted material.

2 Theoretical background: researching the social dimension of variation and change

> [Lawyer:] Now give an account of what happened.
> (OBC, t17450530-17)

This chapter outlines the theoretical background of the study and lays important groundwork for the following analysis. Key assumptions and aims of the historical sociolinguistic approach and the importance of social factors for variation and change are discussed in 2.1. Section 2.2 outlines the challenges involved in accessing the language of the past, which are a prime concern for any historical study. Section 2.3 introduces Late Modern England with reference to the social context, including language attitudes, and key linguistic developments of the period. Knowledge of Late Modern society is crucial to assess which of the issues addressed in 2.2 are especially relevant for the present study and thus informs the methodological decisions made in Chapter 3. Section 2.4 summarizes the chapter.

2.1 Historical sociolinguistics: development, key assumptions and aims

The present study is situated in the field of historical sociolinguistics, which emerged as a subdiscipline of its own in the 1980s. Of course, historical inquiry into language which integrated social aspects of language use had been conducted long before then, but it was not until a little over 30 years ago that an attempt was made to apply the models and methods of modern sociolinguistics (going back to Weinreich et al. 1968 and Labov's subsequent work, e.g. Labov 1972a) to historical data, notably with studies like Romaine (1982a) and Tieken-Boon van Ostade (1987).

The main objective of historical sociolinguistics is summarised in a first handbook article on the subject, featured in Ammon et al. (1988):

> to investigate and provide an account of the forms and uses in which variation may manifest itself in a given speech community over time, and of how particular functions, uses and kinds of variation develop within particular languages, speech

> communities, social groups, networks and individuals
> (Romaine 1988: 1453)

The latest edition of the handbook retains the exact phrasing (Romaine 2005), which reinforces the continuing validity of these aims. The study of language in a historical sociolinguistic perspective can either focus on socially motivated variation at a given point in the past or socially motivated change throughout time (Bergs 2005: 12).

The results of historical sociolinguistic research have not only provided valuable insights on the social component of language variation and change in the past but also contextualised and furthered our understanding of results gained from sociolinguistic work in contemporary settings. Historical sociolinguistics has helped to "evaluate and re-assess" some of the findings of modern sociolinguistic research (Kytö 2011: 431). For instance, research carried out on the Corpus of Early English Correspondence (CEEC, Nevalainen et al. 1998) and its extended version (CEECE, Nevalainen et al. n.d.) showed that the curvilinear hypothesis, which was first established on the basis of contemporary data, can also be also applied to Early Modern data (Britain 2012: 460, referring especially to Raumolin-Brunberg's (2006: 131) discussion of changes from subject pronoun *ye* to *you* and 2SG inflectional *-th* to *-s*, which begin among "the middle-ranking people").

In order to investigate socially motivated variation and change, historical sociolinguistics essentially applies synchronic sociolinguistic methodology to past contexts. The basis for this is the Principle of Uniformatarianism, which states the following:

> The linguistic forces which operate today and are observable around us are not unlike those which have applied in the past. [...] sociolinguistically speaking, it means that there is no reason for believing that language did not vary in the same patterned ways in the past as it has been observed to do today.
> (Romaine 1982b: 295)

Originally introduced in geology in the 18th century (Bergs 2012: 81), this notion of uniformitarianism has since been made use of as an explanatory device by historical linguists (Lass 1997: 25–29) and sociolinguists (Labov 1972b: 275). It is the theoretical foundation that allows studying and theorising about the language of the past in the first place. Of course, subscribing to the notion that language in the past varied "in the same patterned ways" (Romaine 1982b: 295) as nowadays does not give

us license to assume that sociolinguistic patterns identified for present-day English communities (e.g. the association of prestigious forms with women's speech; Labov 2001: 266) can be applied to the past in a one-to-one relationship. It means, however, that we can find sociolinguistic patterns for past stages of the language as well.

These patterns will be associated with the values and norms of the time but ultimately reflect people's basic needs. Meyerhoff (2001: 63–71) lists the following motivations for sociolinguistic variability, as put forward by research in present-day communities:

- accruing social capital ('accentuate the positive')
- avoiding or minimising risk ('eliminate the negative')
- maximising fit, yet maintaining individual distinctiveness ('the balancing act')
- testing hypotheses about other speakers ('it's a jungle out there')

People in the past would have had the same basic needs in terms of belonging to groups and constructing (facets of) their identities. As these processes rely on language to a large extent, it is a reasonable assumption that people in the past also shared these motivations for sociolinguistic variability with the speakers of today. What must be assumed to be variable, though, is who wished to align themselves with whom and what was considered valuable (or 'positive') or risky ('negative') in terms of linguistic behaviour in a given situation for a given person. The fact remains that past societies are different from present-day ones, and different groupings and dimensions were relevant for their members.

The task of the historical sociolinguist, then, is to "try to discover *how* different the past was" (Nevalainen & Raumolin-Brunberg 2012: 26). This involves reconstructing past stages of both language and society. Getting a solid grasp of the system of values, beliefs and relevant groupings for past societies is of the utmost importance to avoid anachronism (Bergs 2012) and to come to meaningful conclusions about language use in these societies. This reconstruction effort, the challenges of which are discussed in detail in 2.3, is only possible with recourse to other disciplines, especially (social) history. In fact, historical sociolinguistics can be characterised as a "hybrid" (Bergs 2005: 8–9) or "interdisciplinary" (Nevalainen & Raumolin-Brunberg 2003: 8) venture at the intersection of linguistics, social sciences and history.

Just like synchronic sociolinguistics, historical sociolinguistics is not a uniform entity but made up of a number of different approaches, theoretical assumptions and foci. Broadly, four major paradigms are distinguished in contemporary sociolinguistics (see Table 2): sociology of language, variationist sociolinguistics, interactional sociolinguistics and the ethnography of communication.

paradigm/ dimension	**sociology of language**	**variationist sociolinguistics**	**interactional sociolinguistics**	**ethnography of communication**
informed by	sociology	dialectology, historical linguistics	discourse studies	anthropology
object of study	status and function of languages and language varieties in language communities	variation in grammar and phonology; linguistic variation in discourse; speaker attitudes	interactive construction and organisation of discourse	patterned ways of speaking, sociolinguistic styles / registers
describing	norms and patterns of language use in domain-specific conditions	the linguistic system in relation to external factors	organisation of discourse as social interaction	situated uses of verbal, para- and nonverbal means of communication
explaining	differences of and changes in status and function of languages and language varieties	social dynamics of language varieties in speech communities	communicative competence; verbal and nonverbal input in goal-oriented interaction	functional appropriateness of communicative behaviour in various social contexts

Table 2. Sociolinguistic paradigms (based on Nevalainen & Raumolin-Brunberg 2012 and Dittmar 1997: 99–100)

They form a continuum between macro-sociolinguistics (focussing on the sociolinguistics of society and issues like multilingualism, language policy and standardisation) and micro-sociolinguistics (focussing on social interaction in language use) (Nevalainen & Raumolin-Brunberg 2012: 30).

In principle, all four paradigms can inform historical research, but their applicability is constrained by our knowledge of historical detail and the quantity and quality of the available data (Nevalainen & Raumolin-Brunberg 2012: 32). While the sociology of language is relatively easily extended to past contexts (Nevalainen & Raumolin-Brunberg 2012: 30), the other three approaches are more difficult to apply to the past. Interactional and ethnographic research needs access to spoken language and

para- and non-verbal information, which is not directly available to the historical researcher. Historical studies applying such methods therefore usually require considerable workarounds (Nevalainen & Raumolin-Brunberg 2012: 32). The more quantitatively informed variationist approach is dependent on sufficient linguistic data by a variety of speakers (Nevalainen & Raumolin-Brunberg 2012: 32), which can be a challenge when working with historical texts.

The relationship between macro- and micro-approaches is best seen as mutually beneficial and complementary. Two remarks on the matter shall suffice to illustrate this point: Commenting on the relationship between variationist sociolinguistics and the sociology of language, McColl Millar (2012: 41), for instance, states that "without the knowledge of one sub-field, it is difficult to talk intelligently about the other", and Auer & Voeste (2012: 267) explicitly warn against discussing variables on the micro-level in a quantitative perspective without adequately integrating the larger social context, i.e. the macro-perspective, into historical studies.

The present study uses a primarily variationist[3] approach: based on electronic corpora, linguistic variables are investigated in terms of their correlation with extralinguistic variables. Mindful of warnings such as the above-mentioned one by Auer & Voeste (2012), particular care will be taken to integrate the larger historical context into the analysis. This is done, among other things, by identifying social dimensions that are meaningful for social organisation in the Late Modern period and by critically reflecting on the production process of the trial proceedings, which represent the basis for the analysis. Speakers' linguistic choices can only be sensibly discussed when we know in what context they made these choices: for the present study, larger societal issues (such as social mobility, language attitudes, notions of propriety and politeness for speakers of different social backgrounds) as well as issues particular to the source texts (language in court, print production) play a role. The analytical procedure is discussed in detail in Chapter 3.

Variation and change are of course very complex issues that cannot be captured by unidimensional explanations. From a sociolinguistic perspective, a large part of this complexity is due to the fact that variation and change are processes originating with individual speakers – complex human beings living in a particular time and place. It is

[3] The terms 'quantitative sociolinguistics' or 'correlational sociolinguistics' are also used in the relevant literature.

no wonder that a variety of different approaches and explanatory devices are employed in historical sociolinguistics, depending on the respective context. What all historical sociolinguistic research has in common, though, is a "theoretical perspective which assigns preference to explanations based on the agency of speakers (or group of speakers) rather than abstract linguistic systems or universal cognitive mechanisms" (Deumert 2003: 19). Linguistic change, for example, is considered a result of speakers' desires to mark, among other things, social identities or stylistic difference (Milroy 1992a: 86).

This does not deny that systemic or cognitive factors also play a role. Acknowledging the multitude of factors involved in language change, Labov's three-volume work *Principles of Linguistic Change* surveys internal factors (Labov 1994) and cognitive and cultural factors (Labov 2010) next to social factors (Labov 2001). The way in which these different influences work together in language change is explained as follows by Hickey (2012: 403): the structural properties of languages and cross-linguistic developmental preferences, ultimately rooted in the mechanisms of language production and processing, provide the framework for the "linguistic course" of any change. It is due to social factors, though, that changes are initiated in the first place, as "social factors determine whether variation, inherent in all languages, is carried over a threshold, after which it becomes change in the community in question" (Hickey 2012: 403).[4] In a similar vein, Chambers (2013: 318) argues that, at a given point in time, "[t]he linguistic conditions may be sufficient, but it is the social conditions that are deterministic" for a change to take place. Linguistic change can thus rightfully be called "a social phenomenon" (Milroy 1992a: 86).

'Social factors' and 'social conditions' in the sense of the above-mentioned quotes can be very diverse, ranging from contact between speech communities over language policy and prescription to individuals' identity construction and relationships. Interestingly, social factors may 'interfere' with systemic factors once a change is underway and e.g. affect an emerging patterning that would lean towards symmetry based on systemic factors (Hickey 2012: 403). In the Dublin Vowel Shift during the late 20th century, for instance, the general retraction of low back vowels did not take

[4] The only exception to this rule is presented by change in early childhood, which, Hickey (2012: 403) argues, is "internal and system-driven and definitely free of external motivation".

place when /a/ occurred before /r/ because the retraction of /a/ in this context was stigmatised in the community, Hickey (2012: 405) shows.

The overall importance of social factors has been consistently highlighted by past research. However, this research has also shown that there is much yet to explore:

> Empirical findings - over the last fifty years, in particular - demand that we assume that social historical (or, if you prefer, sociolinguistic) forces underlie (or interact with) all language change. Sometimes, however, we have limited evidence for what these interactions might have been. (McColl Millar 2012: 42)

Shedding light on these interactions mentioned by McCollar Millar and on variation and change in their social context in general is the job of historical sociolinguistics. Not all limitations can be expected to be overcome, but we can certainly work on refining the picture.

2.2 Accessing the language of the past: challenges and strategies

Having mentioned that reconstructing both the language of the past and the past state of a society falls into the domain of historical sociolinguistics, it is necessary to discuss what this entails. There is certainly no shortage of potential pitfalls as "historical sociolinguists ply their trade in non-optimal conditions", as Britain (2012: 456) puts it. In particular, there is concern about the representativeness of historical data as it is stylistically and socially unbalanced: it is skewed in favour of written styles and the language of the upper echelons of society. The present section therefore addresses possibilities of adequately dealing with the social imbalance of one's data (2.2.1), with the restriction to written sources in historical sociolinguistics (2.2.2) and with studying texts from diverse and changing genres (2.2.3).

2.2.1 Social imbalance in historical sources: Whose language survived?

The historical (written) record does not equally represent all members of a given society. In fact, it is made up of "those written sources that survived for long enough to be consulted": due to social inequalities, this leaves us with a body of evidence created principally by higher-ranking men for large parts of history (Cameron 2008: 293). The sources at our disposal thus under-represent or – in the worst case – even erase the

"socially backgrounded" (Kiełkiewicz-Janowiak 2012: 325) such as women and the lower social classes. These backgrounded groups were less likely to be in a position to produce written linguistic material, especially outside of the private sphere. In addition, they were less likely to be in focus in written material produced by the privileged.

Even for the more recent past, these issues need to be taken into account. The Late Modern material at our disposal in its various forms (letters, trial transcripts, novels, etc.) also mostly reflects a tiny privileged part of society. Reasons for this include the comparative lack of status, financial means, legal rights and access to education (and thus literacy) of large parts of the population. In the beginning of the period (around 1700), the issue is more pronounced than towards the end (1900s), but it persists throughout.

In general, literacy in Late Modern England was much lower than today and distributed unequally among population groups (Cressy 1980: 177). Signature literacy (i.e. the ability to sign one's name) was greater among men than among women. It is estimated that around 1700 ca. 25% of women and 40% of men had signature literacy in England. These figures rise to 40% for women and 60% for men around 1800. The difference between the genders in this area did not level off until the early 20th century. Between 1850 and 1911, male signature literacy rose from just under 70% to 99%, and female signature literacy rose from 55% to 99% between 1850 and 1913 (Cressy 1980: 177–178). Factors like high social standing (see Cressy 1980: 118 for Early Modern data) and proximity to urban centres were beneficial for literacy. Literacy rates in London were for instance higher than in rural areas: while the overall literacy rate of women in England was 25% around 1700, women living in London and its suburbs had already attained a literacy rate of 50% at the time thanks to an "educational revolution" in the metropolis (Cressy 1980: 147).

While differences in literacy surely led to fewer texts by women and people of lower rank being produced, other factors also played a role, such as the type of text a person was likely to produce. When socially disadvantaged groups produced writing, their texts were more likely to be confined to the private sphere and thus less likely to survive over decades and centuries. There are notable exceptions like Martha Ballard's

diary,[5] but printed texts with multiple copies generally had a much better chance of being preserved. However, it should be noted that the likelihood of ever producing a printed text was considerably tied to social status, gender and financial means. In 17th century England, for instance, only 2% of printed texts were authored by women (Cameron 2008: 294).

This imbalance of the historical record is keenly felt especially in quantitative analyses that rely on a sufficiently large sample size to make meaningful comparisons, e.g. across genders or social groups. In their investigation of morphosyntactic features in Early Modern English correspondence in the CEEC, Nevalainen & Raumolin-Brunberg (2003: 137) are forced to discuss the factor social stratification solely based on data by men because they have almost no letters by lower-ranking women at their disposal. Consequently, this eliminated the possibility of investigating different social classes among women writers, who are already underrepresented in the corpus in the first place. These issues are less pronounced in the Old Bailey Corpus, but nevertheless present (see 3.2.4). It will therefore be important to find strategies to deal with these imbalances.

2.2.2 Stylistic imbalance in historical sources: Insight into spoken language?

Modern sociolinguists formulated a preference for analysing the 'vernacular', a speaker's most relaxed and informal spoken style (Chambers 2009: 4–5). Some scholars argue that this should also be the focus of historical sociolinguistics so as to have "a common field of reference" with contemporary sociolinguistics (Tieken-Boon van Ostade 2000: 442). As difficult as it is to access a contemporary speaker's vernacular (e.g. without having the fact that they are being recorded interfere with their speech), the lack of recordings of spoken language for large parts of history makes it impossible for historical linguists to access vernacular data in the above-mentioned sense. To come to terms with this dilemma, two basic strategies have emerged among historical sociolinguists: looking for the 'next best thing' to vernacular speech or moving past the focus on the vernacular.

[5] Martha Ballard (1735-1812) was a midwife in New England and kept a diary beginning in 1785, which is available online (Film Study Center at Harvard University 2000).

Solution number one, i.e. finding “maximally speech-like data” or “a surrogate vernacular”, which was hoped to illuminate processes of change originating in the spoken language, was an important topic of discussion already in the early stages of historical sociolinguistics in the 1980s (Nevalainen & Raumolin-Brunberg 2012: 23, pointing to studies like Kytö & Rissanen 1983). Today, the view that a text should be as close to speech as possible to be useful in variationist analysis is still frequently found (Montgomery 1997: 227, Schneider 2013). To a certain extent, the Old Bailey Corpus is part of this tradition of inching as closely as possible towards the vernacular of a given period. As its subtitle “A corpus of spoken English” indicates, the OBC was at least in part developed with the idea in mind of creating a repository of speech-related data.

There are many different recommendations to be found in the literature on which written genres are most suitable to gaining insight about the spoken language of the past. Based on their long years of experience with the CEEC, Nevalainen & Raumolin-Brunberg (2012: 32) argue that “personal correspondence provides the 'next best thing' to authentic spoken language”. Other recommended texts include trial proceedings (Hope 1993, Tieken-Boon van Ostade 2000: 446, Baker 2010: 77–78), diaries (Elspaß 2012: 165) or drama (Baker 2010: 77).

Yet, it remains a fact that speaking and writing are different processes with different outputs, and that written language is generally more conservative and formal than spoken language and thus exhibits a lower potential for variation (see Hernández Campoy & Schilling 2012: 68, also for further references). This applies even to speech-related writing: simply putting words to paper (or other materials) reduces the potential for variability and innovation because writing is “a self-conscious and monitored activity” (Bergs 2005: 19). Bergs (2005: 19–20) thus rejects the idea of a written vernacular in the sense of a totally unmonitored style. Instead, different written styles should be considered more or less self-conscious styles, conceptualised “on a straight line without an endpoint” (Bergs 2005: 19–20).

This leads to the second option for dealing with the inaccessibility of the vernacular in written texts: abandoning the focus on the vernacular. There are at least two very good reasons for this, the first being that ‘the vernacular’ is more an ideal

than a real speech style that can be observed in the first place,[6] and the second being that interesting and relevant linguistic variation is not restricted to the vernacular, however defined. Variation, while perhaps reduced, is found in even in the most elaborate style (Bergs 2005: 21), which means there is something worth researching for sociolinguists in any register[7] or style[8] of written and spoken language: "[f]or the sociolinguist, no matter whether concerned with present-day or historical data, *any kind of variation will do*" (Bergs 2005: 18).

Several scholars have made the point that historical sociolinguistic studies do not draw their legitimacy from the aim of analysing the vernacular. Bergs (2005: 20) stresses that "there is no built-in need to hypothesise about the spoken vernacular for any historical period". While it is an interesting idea that written sources can serve to reconstruct speech or at least certain features of speech, he warns that "historical sociolinguists should take pains to avoid manoeuvring themselves into a position where the hunt for the spoken vernacular takes precedence over written evidence" (Bergs 2005: 20). Similarly, Finegan & Biber (2001: 239) point out that writing is "deserving of sociolinguistic study not only as a potentially significant interactant with the forms of spoken language but also as a major mode of discourse worthy of sociolinguistic analysis in its own right" (see also Romaine 1982b: 295 or Tieken-Boon van Ostade 2000: 456 for similar arguments).

In the present study, I adopt the position that the transcripts of courtroom interaction are indeed worthy of study in their own right, as a particular genre created in a particular context. But I also assume that there are hints to the spoken language of the period to be discovered in them, which is why I will also explore this aspect. For this to be successful, it is necessary to develop a firm understanding of the ways in which different written texts may reflect speech. In their excellent exploration of

---

[6] Bergs (2005: 17), referencing an argumentation already put forward in Milroy (1992b: 66), remarks on the idealised nature of 'the vernacular' and the difficulty of eliciting so-called vernacular style: there is no way of establishing when someone is at their most relaxed, and any attempt to elicit the vernacular may interfere with a person's speech monitoring. Instead of aiming for an elusive vernacular, linguists should consider real-life speech and writing as composed of many structured varieties that are more or less close to the idealised norm (in Coseriu's sense) or the idealised standard language. Depending on the distance to the idealised norm/standard, varieties and styles can be defined.

[7] The term 'register' can be applied to both spoken and written language. Registers "exhibit a certain cohesion in terms of possible interaction types, aims, and contents, producing lexico-grammatical similarities on a more general level than text types" (Claridge 2012: 238–239): for instance, the domain of law constitutes the legal register with its various genres (e.g. laws, depositions, trial transcripts) and text types (differentiated based on linguistic characteristics rather than their function and typical structure, as is done for genres).

[8] Very generally, style is the situationally distinctive use of language, whereby different styles are mainly associated with different degrees of formality (Crystal 2008: 459–460).

speech-related genres in the Early Modern period, Culpeper & Kytö (2010: 2–3) remind us that in order to assess what spoken face-to-face interaction in past periods was like, we first have to answer the question what written texts representing spoken face-to-face interaction were like. Fortunately, there is very thorough work discussing the properties of historical speech-related texts that I can build on. The remainder of this section summarizes the relevant discussions in Schneider (2013)[9] and Culpeper & Kytö (2010), which inform my further procedure.

At the outset of his evaluation of different speech-related text types, Schneider (2013: 58) explains that the reconstruction of a speech event from a written record always involves 'removing a filter': any written record of a speech event is conceptualised as "a filter between the words as spoken and the analyst". According to the 'Principle of Filter Removal' that he introduces, it is a historical linguist's job to remove this filter:

> a primary task will be […] to 'remove the filter' as far as possible, that is, to assess the nature of the recording process in all possible and relevant ways and to evaluate and take into account its likely impact on the relationship between the speech event and the record, to reconstruct the speech event itself, as accurately as possible. (Schneider 2013: 58)

Depending on what kind of written record one is dealing with, the filter may be composed of various layers (Schneider 2013: 60–61). Based on this idea, Schneider (2013: 61) distinguishes five text categories that represent a "continuum of increasing distance between an original speech event and its written record" based on the reality of the speech event, the relationship between the speaker and person who recorded the utterance and the temporal distance between speech event and time of recording: 1) recorded, 2) recalled, 3) imagined, 4) observed, 5) invented. Among these, recorded texts, such as trial transcripts, are "the most reliable and potentially the most interesting", Schneider (2013: 62) argues, "provided that they are faithful to the spoken word and the speech thus recorded represents the vernacular".

This prerequisite of being a faithful representation of the spoken utterance, labelled validity, is a rather tall order for any written text. One obvious problem is the

[9] There is an earlier version of this handbook article, namely Schneider (2002) that already mentions many salient points. The 2013 article is an updated version, which adds "New Perspectives", as its subtitle indicates (Schneider 2013: 57).

fact that English phoneme-grapheme correspondences do not depict all differences in pronunciation (Schneider 2013: 73). In the case of the Old Bailey Corpus, other issues, such as the shorthand system and its ability to represent variation in speech, also need to be considered (see Section 3.2). On top of the issue of validity, there is also the issue of representativeness, i.e. to what extent the linguistic behaviour in a given text is indicative of the linguistic behaviour of the speech community as a whole (Schneider 2013: 68–71). Obviously, the conclusions of any study will depend heavily on how representative a given text can be assumed to be, and who or what it represents in terms of social groupings, style, genre, etc.

Another in-depth discussion of speech-related genres is found in Culpeper & Kytö (2010), where three different types of such genres are distinguished: speech-like, speech-based and speech-purposed. Speech-based genres comprise texts that are based on speech events in the real world, such as trial proceedings (Culpeper & Kytö 2010: 17). An important caveat applies: most historical speech-based texts are actually reconstructions based on notes, not comparable to recordings, with no guarantee that speech events were recorded accurately (Culpeper & Kytö 2010: 17). Speech-like genres, like personal correspondence, are characterised by features of communicative immediacy (Koch & Oesterreicher 1985), which is primarily a hallmark of the spoken medium but not restricted to it. Features of communicative immediacy include context embeddedness, deictic immediacy, dialogue, communicative cooperation and spontaneity (see Koch & Oesterreicher 1985: 19–21, Culpeper & Kytö 2010: 11).[10] Speech-purposed genres, finally, are "designed to be articulated orally" and may present either monologues, e.g. sermons, or dialogues, e.g. plays (Culpeper & Kytö 2010: 17).

The relationships between these categories are shown in Figure 2. It becomes obvious that the categories overlap to some extent, and that some genres or at least some texts in particular genres may be members of several categories at the same time: plays, for instance, are both speech-like and speech-purposed, as they integrate features of the language of immediacy (making them speech-like) and are meant to be performed (rendering them speech-purposed).

[10] The opposite of the language of communicative immediacy is the language of communicative distance. It is characterised by the opposite of the features associated with immediacy, such as low spontaneity and low context embeddedness. A brief summary of the relevant points from Koch & Oesterreicher (1985) is found on pages 10-12 in Culpeper & Kytö (2010).

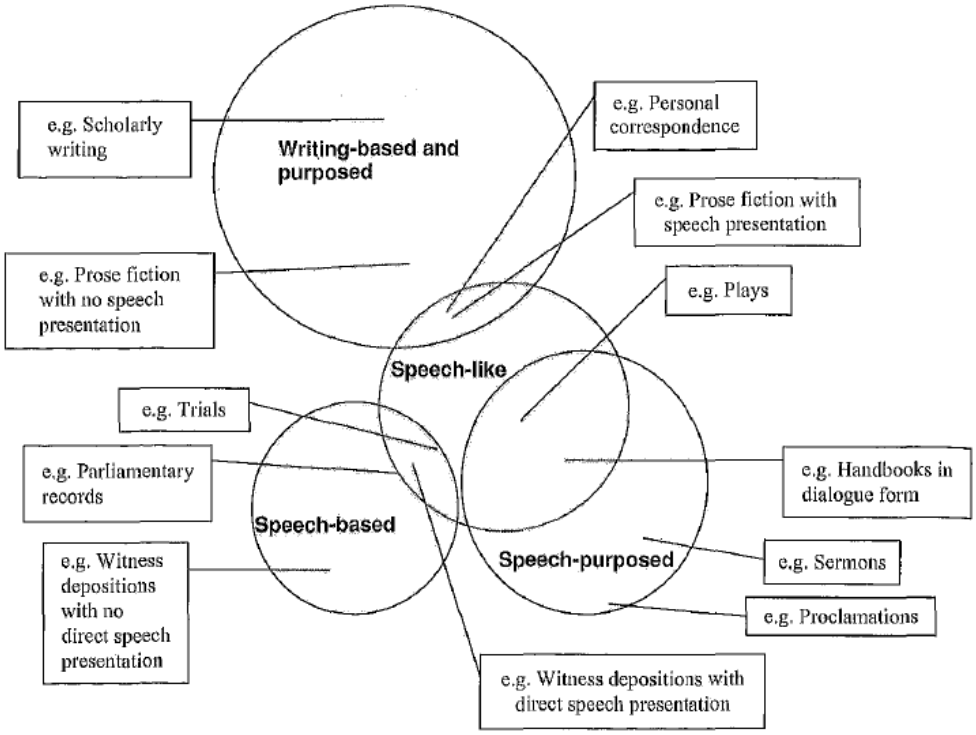


Figure 2. Interrelations between written speech-like, speech-based and speech-purposed genres, as well as writing-based and purposed genres (Culpeper & Kytö 2010: 18)

In the end, all speech-related genres remain "approximation[s] to spoken language of various kinds and in various degrees" (Culpeper & Kytö 2010: 3), with their unique advantages and drawbacks in linguistic analysis. Consequently, Culpeper & Kytö (2010: 3) recommend applying a "triangulation" procedure: they argue that studying several different speech-related genres makes it possible "to reconstruct with some confidence what real spoken dialogue was like". The importance of placing results from one corpus or text in context is also emphasised by Schneider (2013: 73): If the results of a linguistic study show 'external fit', i.e. agree with findings of other studies or reflect familiar linguistic distributions, it is likely that this conformity is caused by both sources depicting the same reality.

On a more general level, the question exists what the overall relation between spoken and written language was at a given point in the past. Romaine (1982b: 295) argues that any historical sociolinguistic theory must include a "sophisticated and coherent account of the relation between spoken and written language". Other historical linguists agree: Smith (1996: 15) calls the "clarification" of the relationship between speech and writing the "most important act of evidential contextualisation

needed in an historical study of English". This is quite a challenge as the relation between written and spoken language is subject to change cross-culturally and throughout history (Romaine 1982b: 295).

How difficult it is to make any statements about this relation is addressed by Smitterberg (2008), who argues that, in the end, the nature of the speech-writing-relation at any point before the advent of audio recordings must remain a matter of speculation. While the author is able to chart trends like the gradual introduction of characteristically conversational features like progressives and phrasal verbs into 19th century writing and proposes colloquialisation[11] as an explanation, he acknowledges that this does not provide direct information on the spoken language of the period (Smitterberg 2008: 286). In fact, several different underlying causes could be behind the growing incidence of conversational features in writing:

> It may be the case that both speech and some written genres changed towards a more frequent use of features that were already common in speech, so that the quantitative distance between speech and informal writing was maintained while the distance between informal and formal writing grew. It is also possible that the frequency of these features was not markedly higher in speech than in informal writing.
> (Smitterberg 2008: 286)

This difficulty provides another reason to clearly differentiate between speech and speech-related written language: we can describe the latter while we are limited to theorising about the former.

Nevertheless, efforts have been made to model the relationship between spoken and written language to give researchers a place to start when hypothesising about historical speech. Krug (2000), for instance, presents a model of change in spoken and written genres based on his study of emerging modals throughout the history of English. Very roughly summarised, the model captures changes from below in different types of spoken and written language as a set of S-curves, of which the one representing informal spoken language is the most progressive, as it represents the genre in which innovations originate (see Figure 3):

[11] The term 'colloquialisation', coined in Mair (1997), refers to linguistic features commonly associated with conversational speech rather than writing becoming established to a greater extent than previously in some written genres (Hundt & Mair 1999: 225–226).

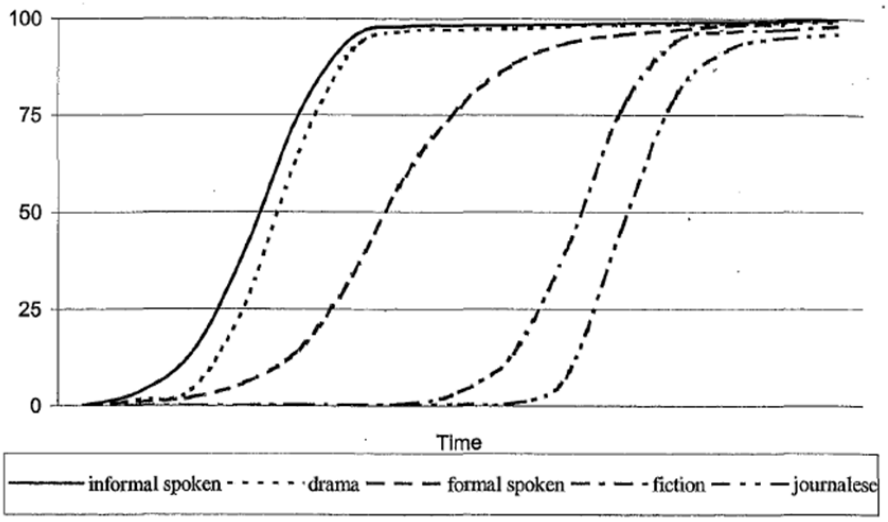


Figure 3. Innovation diffusion in spoken and written language, as presented in Krug (2000: 196)

In the other genres depicted (drama, formal spoken, fiction and journalese), the slopes of the curves are different. Krug (2000: 197) explains that "the gradient will be steeper for genres that are highly susceptible to change (e.g. informal spoken, drama) and gentler for more rigidly codified genres". Frequentative pressures are exerted by varieties that are ahead in a given change. For changes from below, the following scenario, as captured by the S-curves in Figure 3, is envisioned:

> Informal face-to-face conversation generally initiated the changes. This exerted pressures on more formal spoken varieties. Spoken language was then seen to trigger changes in written English genres, whose susceptibility to change depends primarily on their respective degrees of codification.
> (Krug 2000: 194)

In the above model, dramatic texts are considered especially close to spoken language, while journalese is most resistant to innovation. Formal spoken language, a category that can be assumed to include sermons or the spoken passages in trial proceedings, occupies a middle ground between these extremes. However, Krug (2000: 198) draws attention to the fact that genre conventions may change, causing potential intersections between curves. This reinforces the point that the relationships between genres and between speech and writing are not easily captured, especially when a diachronic perspective is introduced.

The way in which people engaged with written English and whether the majority of a speech community was capable of producing it are probably factors that affected the relationship between speech and writing. For Early Modern English, for example, Görlach (1991: 12) argues that speech and writing "were further apart than they are in modern speech communities" because competence in written English was less widespread and there was less need to frequently switch between spoken and written usage. For Late Modern English, it would then be reasonable to presume that the gap between speech and writing was narrower than in Early Modern English, but not as narrow as in present-day English. However, this formulation is an oversimplification as there is no uniform 'written language' that developed in a particular way in relation to 'the spoken language'. Instead, different classes of texts behave in different ways and develop along different trajectories.

2.2.3 Genres and genre conventions between spoken and written norms

In the end, the study of writing requires the same fundamental skills that are needed when working with recorded speech, i.e. "a need for the qualification, assessment and interpretation of one's sources", Schneider (2013: 77) observes. On the same page, he qualifies his statement, saying that working with written sources requires "somewhat more judgment and assessment" (Schneider 2013: 77). This is especially true in a historical dimension, where contextual information on a piece of writing, its purpose, authors and audiences is not readily available but needs to be uncovered.

A helpful tool in this endeavour of assessing texts is provided by the concept of genre, which is also applied in Culpeper & Kytö (2010)'s typology of speech-related texts introduced in 2.2.2, where distinctions are made between e.g. correspondence, sermons and trial proceedings. Texts belonging to the same genre share functions, situations of use and a typical structure (Biber 1989: 39).[12] Investigating variation and change within one genre is thus a good way of cutting down on extralinguistic variation, such as different situations of use, that could influence linguistic output.

[12] Linguistic output in the written and spoken mode can be categorised in different ways, notably by reference to genre or text type. While genres are differentiated based on non-linguistic criteria, text types are defined based on similarities in the use of co-occurring linguistic features, i.e. linguistic criteria (Biber 1989: 39). Genre and text type categories thus may cut across each other (Biber 1989: 27). In contrast to registers and styles (see fn 7), which are inventories of linguistic devices, both genres and text types represent classes of texts (Claridge 2012: 238).

Genres are not homogenous entities, however. Firstly, texts within one genre only share general functions and structures – there is room for variation. The main purpose of a personal letter could be a cry for help, a congratulatory message or a long complaint about mutual friends. The structure (greeting – body – closing formula) is relatively fixed, though we might find variation here as well, e.g. someone opting not to use closing or opening formulas. Secondly, genre conventions are not unchanging through time. An Old English will is quite different from a present-day will, for instance.

A lot of work on changing genre conventions has been done based on Biber's (1988) dimensions of variation in English, established for 23 contemporary spoken and written genres on the basis of a factor analysis involving 67 functionally important linguistic characteristics.[13] Several studies apply these dimensions to historical texts (such as Biber & Finegan 1987, 1989), while others conduct a separate factor analysis for texts at a historical stage to generate the stylistic dimensions typical for a given period (Biber 2001). For the development of English genres (termed 'registers' in some of the publications mentioned in this paragraph) in the Late Modern period, both types of study yield complementary results:

> […] over the last three centuries there has been a large expansion in the range of variation among written registers. At one extreme, expository registers have become extremely 'literate', unlike any spoken or written register in earlier periods; at the other extreme, personal written registers have evolved to become much more like spoken conversation.
> (Biber 2001: 106)

Biber (2001: 106) argues that there was a larger overall gap between spoken and written genres in the early Late Modern period than there is today, but that this gap has closed for some genres and widened for others. It has been shown, for instance, that medical prose (considered non-personal and expository) and drama (considered personal) used to be very close to each other on both the involvement dimension (dimension 1 in Biber 1988) and the impersonal style dimension (5) in the 17th century, but developed apart considerably up to the present day (Biber et al. 1998: 211). The fact that what characterizes a given text type or genre is historically in flux sometimes

[13] For details on the methodology see Biber (1988).

makes it difficult to distinguish language change from change in genre conventions or changing linguistic characteristics of text-types (Kytö 2012: 1522).

It is also interesting to note that such developments did not always go smoothly, as an earlier study (Biber & Finegan 1987) of the development of fiction, essays and personal letters from the 18th century onwards shows. When genre conventions were challenged and new forms tried out by writers, the range of variation increased within a genre. As soon as new norms grew established after this experimentation phase, variation diminished again (Biber & Finegan 1987: 75). That a large range of variation was found in 18th century prose in particular is traced back to a "considerable conflict concerning literacy" that was unfolding at the time: extremely oral styles, having in mind popular audiences, were found on the one hand, and extremely elaborated, abstract styles, intended for an elite readership, on the other (Biber & Finegan 1989: 514). Eventually, a shift towards more popular literacy was accepted in the 19th century (Biber & Finegan 1989: 514). This also highlights that genres do not 'develop' along these dimensions on their own but that an underlying social negotiation process among language users is shaping texts.

2.3 Accessing the social context of Late Modern England

After discussing the challenges of accessing the language of the past, it is now time to turn to the challenges of approaching past societies that produced the language data under investigation. The further back we go in time, the more difficult it is to reconstruct the particular social context in which utterances were made and texts produced. Although conditions are much better for linguists investigating more recent periods, this is not to say that this reconstruction effort is straightforward for the Late Modern period. Relevant issues will be discussed in the following. Section 2.3.1 tackles major issues involved in any kind of historical study, such as the risk of anachronism. The focus moves to Late Modern English in Section 2.3.2, which outlines important historical events and linguistic developments of the period, including the change in status that language experiences in modern Europe. The rise of a codified standard of English and language prescription are dealt with in the final

section in more detail (2.3.3), also with regard to the impact of prescriptive comments on actual usage.

2.3.1 General issues in reconstructing past societies

As mentioned in 2.1, historical sociolinguistics relies on the Uniformitarian Principle, and at the same time acknowledges that the sociolinguistic patterns, rules and mechanisms in a past society will not necessarily be the same as those observed in contemporary English. Basically, this means that the "actual concepts and functions of class, gender, networks, and, most importantly, norms, standards, and prestige, differ radically in different communities" (Bergs 2012: 96), and in fact cease to be useful as analytical categories and tools once we succumb to the fallacy of treating them as 'abstractions' outside time and space (Auer & Voeste 2012: 265).

That the categories and approaches established in sociolinguistic work on contemporary language may not be applicable to past societies can have a number of reasons. Three recurring issues are discussed in more detail below: 1) a concept or distinction that is valid today was not valid in the past, 2) a concept or distinction in the past is not found as such in the present, 3) the evaluation of a particular concept or (linguistic) form was different in the past. Ignoring these issues leads to anachronism.[14]

As an example of category 1, consider the concept of 'teenager' as a life stage between childhood and adulthood, which is not applicable before the 1950s (see Auer & Voeste 2012: 265). Postulating 'teenager' as a separate age group in a study of medieval letters would thus be committing anachronism. At other times, the issue may not revolve around a particular level within a social category (such as 'teenager' in 'age group') but pertain to the way in which extralinguistic categories as such are conceptualised and defined. The contemporary Western economically-based concept of class, for instance, is not suitable for pre-industrial societies (Kiełkiewicz-Janowiak 2012: 311).

As for category number 2, a woman's marital status may serve to illustrate the point. For New England women in the 18th and 19th century, their marital status made a tremendous difference in all areas of their life, including social contacts, employment,

[14] For a detailed discussion of the dangers of anachronism associated with the observance of the Uniformitarian principle in historical sociolinguistics, see Bergs (2012).

legal rights, mobility, etc. Single women had more time to socialise than their married contemporaries and were allowed to take up waged employment, which gave them more economic independence and greater mobility (Kiełkiewicz-Janowiak 2002: 62–63). Married women were much more confined to the sphere of the home (Kiełkiewicz-Janowiak 2002: 63). While marriage is still an important step in many people's lives, a corresponding shift in a woman's role and opportunities is no longer found in present-day New England, for instance.

An important conclusion can be drawn from the examples put forward for the first two issues: to adequately investigate language use in the past from a quantitative sociolinguistic perspective, it is essential to define social variables and their levels in such a way that they are meaningful within the context of the speech community (and the time period) investigated. This involves a good deal of reconstruction as the meaning of any social variable "has to be recovered from the historical text that is the subject of linguistic analysis, as well as from the background writings of the historical period under study" (Kiełkiewicz-Janowiak 2012: 307). If necessary, the researcher must create new (sub-)categories to accurately reflect social structure, groupings and dividing lines between groups.

Nevalainen & Raumolin-Brunberg's (2003) approach to class divisions in Early Modern England in their studies based on the CEEC represents a good example of establishing a social variable (here: social stratification) with meaningful levels. As pointed out, an economically-based system created for industrial societies is not suitable for a time when inherited power was still of major importance. The researchers therefore explored four systems of social stratification customised for Early Modern society (Nevalainen & Raumolin-Brunberg 2003: 136) and finally used an eight-tier-system for the more fine-grained analyses (royalty, nobility, gentry, clergy, social aspirers, professionals, merchants, other non-gentry) and a four-tier system for broader developments (upper ranks, social aspirers, middle ranks, lower ranks). These stratifications are locally meaningful for the letter writers in the CEEC.

Apart from rethinking social groupings, previous work has also shown that other sociolinguistic concepts may need to be adjusted in a historical dimension so as to fit the realities of the periods under investigation. Present-day measurements of network strength, for instance, are not directly applicable to earlier stages of the history

of English because people's lives were different, e.g. in terms of mobility. As a result, network strength scales for Early Modern English (Bergs 2005) and for the 18th century (Bax 2000) have been proposed. Such tailor-made solutions more accurately reflect people's interactions in the respective periods and thus make it possible to apply social network analysis (see Milroy 1980) to past stages of the language.

To deal with issue 3, i.e. the changing evaluation of concepts and linguistic features, researchers need to ensure that present-day judgements and evaluations are not superimposed on historical data. An important notion in this respect is prestige, i.e. the value attached to linguistic forms. Prestige is "a social, not a linguistic, concept" (Milroy 2012: 572). Whether a linguistic form or variety is associated with high or low prestige depends on the social evaluations that speakers attach to it (Sairio & Palander-Collin 2012: 626). These are in turn heavily influenced by the social evaluations attached to speakers using the form or variety in question (Sairio & Palander-Collin 2012: 626). In addition, prestige patterns change over time as the social, cultural and political circumstances change. This is easily illustrated with reference to the English language as a whole, which enjoyed very little prestige in Europe before the 18th century, but has since considerably grown in importance and prestige (Burke 2004: 115).

Moreover, the social context of past societies is also part of the reason why some 'general' patterns observed for contemporary English are not applicable. As an example, Labov (1990: 213) points out that women's role as promoters of change involving prestige forms, confirmed in many studies of contemporary English, presupposes their access to the relevant prestige norm, which is not necessarily the case for women in the past. The notion of a linguistic standard is also problematic in historical inquiry. While it is true that a codified standard usually enjoys overt prestige (Milroy 2001), it is important not to equate standardness with prestige and to acknowledge that the notion of a standard is not applicable to all speech communities (Sairio & Palander-Collin 2012: 628). There was, for instance, no codified standard to speak of for Middle English, so it would not make sense to link observed linguistic developments to speakers' adherence to the standard language.

To conclude, it is the linguist's responsibility to assess whether a social variable is meaningful (e.g. is gender/age/marital status a socially salient distinction in

the community in question?) and which divisions are meaningful (e.g. does a division between childhood, youth, middle age and old age make sense for this community?) Only an "in-depth social understanding of a given demographic dimension" (Kiełkiewicz-Janowiak 2012: 324) makes it possible to adequately conduct historical sociolinguistic work.

2.3.2 Late Modern England: society and language

Periodisation, as is well known, is a conscious structuring effort by historians (or linguists, for that matter) that relies on a consensus among scholars, not something that is present as such in history (Price 1999: 1, see also Curzan 2012). For Late Modern English, a consensus has emerged among linguists to set the boundaries for this period between roughly 1700 and 1900 (Beal 2012b: 64). Social historians, who think in other kinds of periods, discuss this stretch of time under a variety of different labels. In a British, especially English, context, the period may be subdivided into the Georgian and Victorian eras (based on the reigning monarchs) or discussed with reference to decisive historical events and processes, such as the 'Industrial Revolution' or the 'Enlightenment'.

However, some historians come very close to the linguistic construct of the Late Modern period in their analysis: Price (1999: 2), for example, argues that the period between the late 17th and the late 19th century shows remarkable continuity. Throughout the period, the economic growth witnessed was achieved by the intensification of established methods of production. The so-called 'Industrial Revolution', as such a debated concept among historians,[15] mainly increased the scale of manufacture: "productivity gains were sought through the more extensive use of labor rather than through harnessing the technology of the machine" (Price 1999: 27–28). The social and political power was retained by the landed elites, who thrived because of their involvement in economic innovation and profit maximisation (Price 1999: 294). In the political arena, local structures of government rather than central ones remained the most important (Price 1999: 155–191).

Price (1999: 11) argues that "in the late seventeenth century [...] an architecture of society became visible that was to define the major themes of British society for the

[15] See Price (1999: 19) for an overview of the revisionist literature on the 'Industrial Revolution'.

next two centuries". This system was finally eroded by the end of the 19th century when its established structures were no longer able to cope with contemporary challenges such as the Great Depression of 1867-96 and the changing configuration of international relationships, including increasing economic competition from America and Germany (Price 1999: 336–342).

For reasons of brevity, only some of the major social, economic and political events of the Late Modern period can be mentioned here. Beal (2012b: 66–69), who approaches the historical development from a linguist's point of view, lists the English Enlightenment and the associated scientific progress, technological innovation and industrial growth, urbanisation, advances in transport and communications, as well as social reform, e.g. the introduction of compulsory elementary schooling in 1870, as major processes and milestones in the Late Modern period. All of these had an impact on the English language in one way or another.

Scientific progress and technological innovation spurred industrial growth and urbanisation. Dialect contact was a main linguistic consequence of population movements, notably workers coming to the towns and cities. This contact, which was aided by improvements in transport (like new roads) and communication (like the establishment of the Penny Post in 1840) led to dialect levelling and the formation of new urban dialects (Beal 2012b: 67). As a consequence of the expansion of educational provision, more children than ever before received at least elementary schooling and came into contact with the emerging standard variety of English (Beal 2012b: 67).[16] In general, language was seen in a different light than before. Britain was on its way to becoming a fully literate society and a print culture:

> In the 1660s, most people did not read or write. Language was primarily an oral phenomenon, transmitted from mouth to ear. The most powerful national communications systems were the sermons and speeches and readings that people heard every week in church or chapel. But by 1830 all the (pre-electronic) institutions of literacy were going full steam: dictionaries, magazines, anthologies, advertisements, newspapers, cartoons,

[16] It would be overstating the matter to say that all children received elementary education. Between 1750 and 1850 full-time formal schooling was becoming the standard for men of the privileged groups in society, i.e. the aristocracy and the emergent middle classes, but schooling for the vast majority of children was "intermittent, chancy, partial, often wholly informal or at most semi-formal" (Sutherland 1990: 140). After the Education Acts of 1870 and 1880, compulsory nationwide elementary education for children aged 5-10 was introduced in England. However, this did not mean that all children on the registers were in a position to attend school regularly – or at all (Sutherland 1990: 141–146).

> lending libraries, book reviews, women writers, and feminist tracts. (McIntosh 2008: 228)

This elevated the status and reach of written language considerably. Tax reductions on newspapers in the first half of the 19th century and better printing technology played an important part in the expansion of print media, which helped consolidate the written standard (Auer 2012: 942).

Increased commercial opportunities and better schooling created a society with increased social mobility as "[m]oney began to count almost as much as land ownership" (McIntosh 2008: 231). The middle classes emerged as a separate group, characterised as "ambitious" and "upwardly mobile" (Beal 2012b: 68), though within limits. As mentioned above, traditional elites remained powerful throughout the period (Price 1999: 294) and the middle classes were an "essentially urban phenomenon" only (Hickey 2010a: 9). It is the upwards mobility of the middle classes that is considered intimately connected to the "considerable linguistic self-consciousness" in Britain in the Late Modern period (Finegan 1992: 105). Without traditional indicators of high status such as land or university education, they are said to have been especially preoccupied with correct language use (Smitterberg 2012: 953). This was particularly true of the lower middle classes: when they, too, gained some influence in the 19th century, partly as a result of the expanding service sector (Matthew 2001: 542, Smitterberg 2008: 283), they put great effort into speaking and writing 'correctly' (Beal 2004: 116).

It was already in the Early Modern period that the perception of language in Europe saw an important change: through the first publications on vernacular languages such as dictionaries and grammars, language became a status symbol and a "bearer of social significance" (Auer & Voeste 2012: 258). Numerous treatises on language from the period illustrate people's awareness of linguistic variation and of the social evaluations of languages, dialects and accents (Burke 2004: 15–42). In the English context, this trend only intensified in the Late Modern period, which is famous for being the age of standardisation and codification of the English language (see e.g. Auer 2012, Finegan 2012 and Percy 2012 for recent, succinct overviews). While there had been individual and local standardising tendencies before, it was only in the second half of the 18th century that "massive legislation and prescription" and "the full

appearance of all the attitudinal satellites as the component part of the ideology of standardization" appeared (Stein 1994: 5). The emerging standard, especially in terms of pronunciation, was based on educated south-eastern usage (Smitterberg 2012: 953). This makes perfect sense in light of the fact that "London had been singled out as the home of the 'best' English from at least the sixteenth century" and only increased in importance after the Act of Union in 1707, when it became capital not just of England but of Great Britain (Beal 2010: 26).

What standardisation actually entails in a social dimension is very well summarised by Stein (1994: 7):

> 1. in a process of selection, certain variants are not elected as the correct ones;
> 2. the ones not elected receive a connotation as 'vulgar' and 'dialectal';
> 3. the people using these vulgar or dialectal forms are branded as socially and intellectually inferior.
> (Stein 1994: 7)

In the wake of the standardisation process thus came an "increased tendency to assign social evaluation to variation in pronunciation and grammar" (Smitterberg 2012: 957). As variation is especially noticeable in pronunciation, a person's accent became an important measure of the standardness of their language from the second half of the 18th century onwards (Beal 2010: 23). Those who wished to gain status in society were aware of the "exterior standard of correctness" and tried to follow this norm (Labov 2001: 277).

The English language was by no means static in the Late Modern period, which will also become obvious in the detailed accounts of the morphosyntactic features under investigation in this study (Chapters 4-7). In fact, a great deal of variation and also some important changes can be observed at all levels of language, of which I will only discuss morphology and syntax in the following. It is generally acknowledged that language change in the 18thand 19th centuries involved few categorical losses and gains and mostly revolved around constructions becoming more or less frequent, either generally or in particular registers (Denison 1998: 93 makes this point with reference to syntactic change, Beal 2012b: 69 for morphology and syntax).

In a recent survey of Late Modern English syntax, Aarts et al. (2012) take up this distinction between 'categorical innovations' on the one hand and (much more

numerous) 'statistical and regulatory changes' on the other hand for their overview of the language of the period. Only the grammaticalisation of the GET passive and the emergence of the progressive passive qualify as categorical innovations (Aarts et al. 2012: 870–873). Statistical and regulatory changes, though, are quite numerous: Aarts et al. (2012: 873–882) list the progressive, the decline of BE as a perfect auxiliary, the regulation of periphrastic DO as well as shifts in the complementation system and in the use of relativisers.

The low number of categorical changes may easily give the impression of great stability within the Late Modern period. Finegan (1992: 104), for instance, reports that "the grammatical characteristics of the English language remained comparatively stable between 1700 and 1900" (for similar assessments, see e.g. Rydén 1979, Beal 2004: 66). The comparative similarity with contemporary usage is also often invoked, as in this assessment by Rydén (1984):

> It is true that in many cases the main outlines of present-day usage were established by the 18th century and it is possible that differences between present-day English syntactic usage and that of the 18th century are primarily differences in variant-frequency, style and social stratification — few constructions current around 1780 would be impossible today in an overall perspective.
> (Rydén 1984: 511–512)

What this quote also acknowledges, however, is that both differences between Late Modern and present-day usage as well as variability within LModE exist and offer relevant research opportunities for linguists. In fact, it would be wrong to discard a change as insubstantial simply because it is not categorical. Referring to developments in the tense-aspect system (rise of the progressive, progressive passive, GET passive and HAVE perfect), Anderwald (2012: 29) argues that "[l]anguage change in the 19th century did not consist of minor or peripheral re-adjustments of an otherwise fixed system, but affected some core areas of grammar".

Although the Late Modern period is associated with prescriptive norms and the codification of English (see 2.3.3), it offers many opportunities to study variation. Thanks to the comparatively high rate of survival of textual evidence, linguists have access to a variety of formal and informal registers, which allows valuable insights into people's sociolinguistic competence. Social variation, which is at the heart of the present work, is attested at all levels of usage throughout the period (see Smitterberg

2012). In light of the prevalent ideology and the proliferation of normative grammars at the time, it is necessary to discuss to what extent variability may have been constrained by language prescription. The next section is dedicated to this issue.

2.3.3 Language prescription[17] and language use

The growing number of grammars, dictionaries, pronunciation guides and other instructional publications[18] throughout the Late Modern period is well documented. Information on 18th century grammars can be found e.g. in Sundby et al. (1991), a dictionary with data on 187 titles, or in the Database of Eighteenth-Century English Grammars (ECEG; Yáñez-Bouza & Rodríguez-Gil 2010), which contains information on over 300 works. For the 19th century, Görlach (1998), an annotated printed bibliography of grammars, offers an excellent overview. Another notable source is Anderwald's (n.d.) electronic Collection of Nineteenth-Century Grammars (CNG), surveying over 250 grammar books. All these resources include British and American titles.[19]

The impact of language prescription is the subject of lively discussion. Did prescription influence people's language choices and have an impact on language change? Or did the grammars of an era only reflect processes that were already underway, and as such mirror people's usage? Results from linguistic research thus far are mixed. While some studies can establish a link between language prescriptions and individuals' usage, most scholars take the position that it is best not to overestimate the overall impact of prescriptive norms.

That some individuals did change their usage to reflect the standard of the day emerges in several studies. For example, Sairio (2009: 312–313) shows that Elizabeth Montagu, an 18th-century Bluestocking, modified her language considerably

[17] As the term 'prescriptivism' has negative connotations for many linguists, I follow Leech et al. (2009: 263) in using the arguably more neutral term '(language) prescription' for "any conscious efforts to change the language habits of English speakers (or more often, writers)". In the same vein, I use the adjective 'prescriptive' instead of 'prescriptivist'.

[18] Apart from grammar books, there were many other sources of instruction on 'good' English. Especially in the beginning of the 18th century, when the standardisation process was in its infancy, journalists like Joseph Addison, who advanced the ideology of a standard in the *Spectator* (1711), set the tone (a detailed discussion is found in Fitzmaurice 2000). Later, book reviews were a vehicle for linguistic critique and prescription, e.g. in the *Monthly* (est. 1749) and the *Critical* (est. 1756) (Percy 2012: 1009).

[19] Not all grammars and other writings on language of the period are 'prescriptive' or 'normative'. Beal (2004: 90) states that grammars occupied "different points on a prescriptive-descriptive continuum". Some actually favoured a descriptive approach (see e.g. Straaijer's (2010) discussion of the grammarian Joseph Priestley's (1733-1804) treatment of auxiliary choice with mutative intransitives).

throughout her life as her social standing increased and the normative view of language strengthened: she gradually abandons the heavily stigmatised practice of preposition stranding in favour of pied piping, for instance. Another linguistic choice by Montagu illustrates her awareness of social hierarchy: she never uses contracted verbs in correspondence with people of higher status (Sairio 2009: 312–313). This makes sense in light of the fact that contractions were subject to increasingly proscriptive views during the 18th century. Some grammarians advised explicitly against them in correspondence with superiors (Haugland 1995: 175).

In general, though, it is unclear how many people were exposed to or interested in such issues of language use. There is evidence that middle-class social aspirers were the core market for the impressive number of normative texts on language use published at the time (Beal 2012b: 68). Hickey (2010a: 8) also argues that language was mostly an issue for the middle classes in the 18th century. The poorer population groups could not afford to buy the relevant books[20] and were probably preoccupied with more pressing issues. The aristocracy's interest in language studies, he argues, must be considered "doubtful" as grammars were mostly written by their social inferiors (Hickey 2010a: 8). We have to assume, however, that people outside the middle classes, whether of lower or higher status, were at least indirectly affected by the prevalent discourse on language – especially as this discourse gained momentum in the 19th century.

What constituted 'correct usage' was very much of middle-class origin, though: Finegan (1992: 106) remarks that most grammarians and lexicographers came from the middle classes and that their works thus reflected middle-class norms and values. The linguistic and ideological notion of 'correctness' emerging in the late 18th century can even be considered such a middle-class value: the middle classes "adopted and identified with the linguistic and social behavioral forms of the older aristocracy, and in the process established the use of correctness norms as a linguistic class shibboleth" (Stein 1994: 8). Incidentally, this allowed people in the middle of the social hierarchy, themselves often slighted by their social superiors, to discriminate against the lower

[20] Beal (2004: 116) points out that a number of cheaper, more concise and less theoretical handbooks that were aimed at a wider audience appeared in the 19th century. At that point, the mass reading public had been discovered as an audience for literary works in general (Altick 1998: 274–277).

classes (whose language was considered 'vulgar') or those speaking a noticeably regional variety of English ('dialectal') (Berger 1978: 71–72, Stein 1994: 8).

In fact, prescriptive work was primarily focussed on condemning such vulgar or dialectal usage during the process of standardisation. Normative grammars thus did not specify the standard and outline prestigious variants but rather listed what was condemned as incorrect (Auer & Voeste 2012: 258), using a variety of negative labels such as 'absurd', 'dialectal', 'improper' or 'uncouth' (see Sundby et al. 1991: 39, also for further labels used in 18th-century grammars). Hickey (2010a: 16–17) suggests that this reluctance to describe a standard was at least partly a result of the variation found among speakers that prescriptive authors liked to put forward as models of good linguistic behaviour. In addition to that, grammatical norms and notions of what constituted 'good English' were by no means stable throughout the period (Finegan 1992: 104).

Research in the last 20 years has shown that the effects of language prescription in change should not be overestimated (see Auer 2012 for a review of relevant studies). The study by Auer & González-Díaz (2005) on the inflectional subjunctive *(if this **be** your conviction)* and the double comparison (*more wiser*) represents a good example: they compare the treatment of both constructions in 18th-century British grammars with their actual use at the time, concluding that the explanatory potential of prescriptions is very limited indeed in this case. The inflectional subjunctive, which was in decline when grammarians started complaining about its lack of use, could not be saved by prescriptive efforts: promotion of the structure in grammars merely caused a short-lived and rather moderate revival in the late 18th and early 19th centuries (Auer & González-Díaz 2005: 323).[21] The "social downgrading" of the double comparative was mostly complete by the end of the 17th century, i.e. before it was criticised in grammars (Auer & González-Díaz 2005: 335). Prescriptive efforts thus at most reinforced a process that had been underway for a long time (Auer & González-Díaz 2005: 336).

These two case studies support the idea that language prescription lacks the "dynamic quality" of actively promoting language change, but may have a "retarding

[21] Earlier works also report this development of the subjunctive: Strang (1970: 209) writes about a "decline that has continued to this day, reversed sporadically only by the tendency to hypercorrection in 18c and later teachers and writers"; similar observations are found in Görlach (2001: 122).

influence" on ongoing changes (Hickey 2012: 391). Hickey (2012: 391) argues that prescription may impede a change from reaching completion, creating a situation in which the outgoing variant is retained as a possibility alongside the incoming one. For a brief time, this is what happened with the inflectional subjunctive. A second possible scenario, according to Hickey (2012: 391), is one in which prescriptive efforts "stop a development entirely, or at least exclude it from standard forms of a language." Rydén (1984: 514) credits the doctrine of correctness in the 18th century with putting a stop to levelling tendencies such as the use of *was* in the second person singular (e.g. *you was*, see Chapter 7). Where normative efforts do indeed have an impact on language use, this is usually limited to specific registers and depends on a great deal of institutional backing of the advanced norm. It is only due to a "high degree of institutionalization" (Anderwald 2014a: 13) that the progressive passive declined in newspaper language to the degree it did during the second half of the 20th century: the intervention of copy editors who reformulated passive sentences as active sentences was of the utmost importance for this prescription to succeed.

Prescriptive efforts are usually triggered by an awareness of structures in colloquial speech that are considered undesirable for some reason, so that their goal is to exclude them from more formal registers (Hickey 2012: 391). The Late Modern grammarians' reasons for finding a structure undesirable are diverse and not applied consistently across features and grammar books. In the case of contractions, briefly mentioned in the discussion of Elizabeth Montagu's linguistic choices above, justifications for rejecting them ranged from 'sounding bad in speech' over 'looking bad in writing' to being confusing, too colloquial or unnecessary (Haugland 1995: 172–175). Some of the overarching lines of argumentation in 19th century grammars involve reference to accepted usage in the past, but also to "more abstract concepts like *logic* or *analogy*, *elegance* and ultimately *good manners*" (Anderwald 2012: 47). The influence of Latin grammar writing, its conventions and categorisations is also an important factor (Beal 2004: 107–111).

Among the structures condemned by prescriptive works, there is no direct correlation with the spoken/written dimension, either, as an examination of the progressive and the progressive passive in 19th century grammars shows:

> […] the progressive, judging from its present-day distribution as well as 19th century genres a feature more characteristic of the spoken language, is accepted and even commented on positively, while the progressive passive, mostly found in written language, especially in scientific discourse and in newspapers, is criticized strongly […]. (Anderwald 2012: 45)

Negative comments are also not exclusive to innovations, and not all innovations are necessarily considered bad. Anderwald (2012: 29) identifies a number of features that have a bearing on how linguistic phenomena are treated in Late Modern grammars: their text frequency (rare vs. usual constructions), the rate of change (slow vs. rapid), the stage of a change on the S-curve (incipient vs. almost completed changes), and their overall salience. Based on several case studies, it emerges that faster changes, for instance, are subject to harsher criticism than slower ones (Anderwald 2012: 43).

In the end, I agree with Anderwald (2014a: 15) that we must beware of invoking prescriptive influence as a "catch-all explanation for unexpected phenomena, without any further substantiation" whenever we are at a loss to account for change. Research on the connection of prescription and practice so far indicates that "most prescriptive grammarians neither drove nor kept pace with actual changes" and "did not trigger linguistic norms" (Auer & Voeste 2012: 259). What prescriptive grammarians did achieve was a key position as linguistic authorities, in which capacity they furthered a process of verticalisation (Reichmann 1990), i.e. a move towards a system in which variants of variables no longer coexisted on equal terms but formed a hierarchy in which only one variant was considered correct. This changing idea of language, going hand in hand with a limitation or even elimination of variation, characterises the Late Modern period (Auer & Voeste 2012: 259).

2.4 Summary and outlook: an analytical challenge

This chapter discussed the theoretical background and key aims of historical sociolinguistics: to investigate variation and change in the past from a sociolinguistic perspective, i.e. putting speakers, their choices and social realities first when analysing observed language use and theorising about explanations. Due space was accorded to the various challenges included in this endeavour, such as the scarcity and imbalance of surviving material, which makes reconstructing language and society of the past

difficult, and the shifting conventions of individual genres that need to be separated out from general language change. Furthermore, people's changing ways of life lead to different social constructs and groupings being more relevant in one period than another. Importantly, the chapter also summarised tools and approaches available to the historical sociolinguist to make an analysis work despite the inherent difficulties.

In all discussions of how to manage 'deficits' and 'issues', it is important to keep in mind that historical sociolinguistics may be informed by its 'big sisters', traditional historical linguistics and present-day sociolinguistics, but its success should not be measured against their specific goals (although of course there is overlap). If we do this, we back historical sociolinguistics into a corner. Against the standards set by modern sociolinguistic investigations, for instance, the data and methodology of historical sociolinguistics will always fall short in some respect, e.g. in terms of the set-up of categories, the breadth of personal information on speakers, etc. (Nevalainen & Raumolin-Brunberg 2003: 154). Compared to traditional historical linguistics, historical sociolinguistics has done less in identifying general mechanisms of change but rather focussed on analysing shifting patterns of usage and their social meanings in a given context (Roberge 2012: 381–382). But then, these are not the discipline's primary goals.

A different, arguably more productive, way of engaging with historical sociolinguistics is one that acknowledges existing difficulties but is not weighed down by its perceived shortcomings. Bergs (2005: 21), for instance, emphasizes that the discipline "does not *suffer* from a lack of natural, spoken linguistic data, or social data", and further advises the following:

> […] historical sociolinguistics must be bold enough to loosen its ties with present-day sociolinguistics and traditional historical linguistics, and to develop its own methodologies, aims, and theories. (Bergs 2005: 21)

This is not to say that the rich base of knowledge of these two neighbouring disciplines (and of others) should not be used – on the contrary! Instead, they can complement each other. What is made clear, however, is that historical sociolinguists need adapted methods and procedures that are suitable to their aims and the discipline's research programme. In recent years, considerable headway has been made in this regard, for

instance by tapping new sources of data and by critically examining sociolinguistic concepts in terms of their suitability for investigating past stages of the language.

The quote by Britain (2012: 456) at the beginning of Section 2.2 pointed out the "non-optimal conditions" that historical sociolinguists are exposed to in their work. But then, what discipline can say that their conditions are optimal? What is important is to acknowledge existing difficulties and adequately adapt to them when selecting data and methods and putting forward explanatory constructs (and, in the worst-case scenario, to admit that we don't know (yet) what a particular observation means). The following chapter discusses how this is realised in the present work, outlining data and methodology.

3 Empirical foundations: corpora and methodology

> [William Hudson:]
> And are you sure these are the very words I uttered?
> (OBC, t17931204-54)

Several scholars have pointed out that there is a wealth of material for the Late Modern period, much of which has not been accessed by linguists yet (Smitterberg 2012: 963, Tieken-Boon van Ostade 2000: 446). In part, this is due to the fact that interest in Late Modern English has only become more pronounced after the 1990s (Beal 2012a: 16) and that the development of corpora for the period is thus lagging behind in some regards. Although much textual evidence is preserved, both in manuscript and published form, making this accessible in a format suited to linguistic inquiry is no small undertaking. For the present study, the Old Bailey Corpus and the Corpus of Late Modern English Texts, two corpora that intend to make LModE more accessible to the analyst, are used. This chapter introduces these corpora (3.2-3.3) after outlining the benefit of corpora in historical linguistics in general (3.1). The focus of the discussion is on the OBC, which is the main source of data. It is therefore essential to be aware of its opportunities and drawbacks so that analysis and interpretation have an informed basis. Moreover, the variables under investigation are introduced (3.4) and the analytical procedure is outlined (3.5) before a summary (3.6) concludes the chapter.

3.1 Corpora in historical and sociolinguistic studies

For over 50 years, electronic corpora have been used in linguistics. Focussed on authentic language data and empirical evidence instead of introspection, corpus methodology has continually branched out to be applied to more and more linguistic questions, including those concerning language variation and change and their social components. As for many other linguistic subdisciplines, a “contact zone” with corpus methodology has emerged for sociolinguistics (Mair 2009: 8). Historical corpus

linguistics is also well-established. The benefits of using a corpus method to undertake diachronic sociolinguistic analyses as well as some caveats are presented here.

Using corpora has some clear benefits for historical linguists: to describe general trends in the development of a language, researchers need access to a sufficiently large database, which computerised corpora can provide (Kytö 2011: 420). The availability of a large amount of data also facilitates statistical analysis to uncover correlations between extralinguistic and linguistic variables (Kytö 2011: 420). This is an important aspect of variationist work and thus also ties in well with the sociolinguistic enterprise. In addition, corpus-linguistic methodology has extended the possible research questions that diachronic linguistics can ask, and led to new insights on the process of change. For instance, it offers the possibility to assess the transmission of change in a quantitative way (Kytö 2011: 421). Corpora also allow extracting the shifting frequencies of competing constructions involved in change.

Beal (2012a: 17–20) draws attention to the added value that corpus-linguistic approaches provide for historical studies, citing the case of the English progressive as one example. While it is true that Strang (1970) had already outlined the basic development of the progressive in Late Modern English based on a small hand-compiled corpus, later studies such as Smitterberg (2005), which were informed by corpus methodology and based on larger electronic text collections, were able to significantly refine the picture. They also made additional discoveries, for instance that considerations of genre and style are essential to understanding the expansion of the progressive. Corpora have also enabled a deeper insight into the relation between precept and practice in Late Modern England (Beal 2012a: 22) and have in fact led to a re-evaluation of the role of prescriptive grammarians, which has long been overestimated (see 2.3.3, or Auer 2012).

It also seems that the availability of electronic corpora put certain topics on the agenda of researchers or at least shifted priorities. Generally, the availability of corpora has "energised" linguistic research on LModE (Beal 2012a: 16). From the 1990s onwards, the growing interest in the period can be partly attributed to the improved availability of material to study. More specifically, research on lexical and morphosyntactic change and explorations of sociolinguistic and pragmatic questions have flourished, whereas phonology, long the backbone of historical linguistics, has

taken a backseat – simply because there is no adequate corpus to investigate it (Beal 2012a:13, 22-23).

While corpora have undoubtedly broadened linguistic horizons, it is necessary to be clear about what we can realistically expect from historical corpora and about existing problems (see e.g. Rissanen 1989, Durrell 2015 for treatments of problems specific to historical corpora). First of all, historical corpora can never achieve the balanced design of modern corpora because there are gaps in the historical data, and historical linguists need to work with whatever is preserved (Kytö 2012: 1522). Due to the difficulties of gathering appropriate textual material, historical corpora are usually smaller than corpora of contemporary English. This impacts the possibilities of quantitative and statistical analysis, especially if several factors are considered at the same time (as e.g. in a cross-tabulation of forms by gender and class). A key problem is that "the breakdown across the categories distinguished makes representation dwindle away in the data categories" (Kytö 2012: 1516). This challenge has also given researchers cause to question the traditionally very restrictive definition of a corpus as a balanced, stratified sample of language, suggesting that a more flexible definition also including "more open-ended and unbalanced electronic data sources" is useful in a historical perspective (Kytö 2011: 421). It is also my conviction that the standards of modern corpus linguistics cannot be imposed on historical material. I am equally convinced, however, that compilers need to take measures that offset the drawbacks or at least make them transparent to the end-user.

A general concern in a corpus approach is the need for contextualisation of (quantitative) results. Where corpora from different time periods are compared, it is especially important to make sure that we do not create an over-simplified narrative of change out of our observations (Baker 2010: 79–80) but take differing production contexts, styles, etc. into consideration. Baker (2010: 80) stresses the importance of combining qualitative analysis such as in-depth reading of concordance lines where appropriate, with quantitative analysis to ensure the success of a corpus-linguistic analysis. More generally, corpus data as such do not simplify or improve linguistic research per se, as Durrell (2015: 30) reminds us: "you do have to know what you are looking for" and then interpret what you find "in the light of what else we know about the language in question at the period in question". Data extracted from a corpus only

become meaningful when they are contextualised. For historical corpora, "the support of other disciplines such as historical linguistics, variationist analysis, historical pragmatics, grammaticalisation theory, discourse analysis and statistics" are instrumental in making sense of the collected data, for example observed frequencies (Kytö 2012: 1519).

3.2 Main source of data: The Old Bailey Corpus 1.0

The Old Bailey Corpus (abbreviated OBC, Huber et al. 2012) consists of Late Modern English trial proceedings, which contain ca. 14 million words of recorded speech in the courtroom between 1720 and 1913. As such, the OBC is a collection of speech-related texts recording face-to-face interaction. Having pointed out the necessity to be familiar with one's sources before embarking on a linguistic study (see 2.2), the present section sets out to answer important questions about the OBC: What is the OBC as a corpus like? (3.2.1) What were the *Proceedings of the Old Bailey*, i.e. the basis of the OBC, like as a publication? (3.2.2) What was the recording of spoken interaction in court like? (3.2.3) What was interaction in court like? (3.2.4). Answering these questions will allow a realistic assessment of the material and provide a solid basis for the analysis in the following chapters.

3.2.1 The OBC as a linguistic corpus

The Old Bailey Corpus documents spoken language in the courtroom in 18th and 19th century-England. It is based on the digitised *Proceedings of the Old Bailey*, made available via the project 'The Proceedings of the Old Bailey, 1674-1913', directed by the historians Tim Hitchcock and Robert Shoemaker. As part of the project, the printed *Proceedings* were digitised and integrated into a searchable online database from 2003 onwards. After the initial release, the project has been continually expanded and updated (Hitchcock et al. 2015).[22] It consists of trial proceedings from the Old Bailey, London's central criminal court, which were published between 1674 and 1913.

[22] The citation refers to Version 7.2 of the 'The Proceedings of the Old Bailey, 1674-1913', updated in March 2015. For more information on the project, including historical background information, details on methodological and technical aspects and research based on the *Proceedings*, visit www.oldbaileyonline.org.

While the online database of the project 'The Proceedings of the Old Bailey, 1674-1913' offers a fascinating glimpse into the lives of ordinary people in Late Modern England and allows searches within the trial proceedings, it was not created with linguistic research in mind: the concordance-like output is clumsy for the linguist and cannot be exported (Huber 2007). To make the material in the *Proceedings* accessible for linguistic study, the Old Bailey Corpus (Huber et al. 2012) was compiled at Justus Liebig University Giessen, Germany. OBC version 1.0, totalling ca. 14 million spoken words in 318,000 utterances, was published in June 2013. This version is the basis of this study.[23]

The OBC contains 407 proceedings between 1720 and 1913. All proceedings in the corpus are included in their entirety, but only the spoken passages are counted towards the corpus size of 14 million. The corpus is designed in such a way that the number of spoken words equals roughly 750,000 in all decades, as Table 3 shows.

decade	word count
1720s	72,529
1730s	752,126
1740s	750,450
1750s	746,966
1760s	751,884
1770s	749,475
1780s	765,362
1790s	751,014
1800s	754,809
1810s	769,769
1820s	750,001
1830s	763,903
1840s	765,844
1850s	745,911
1860s	757,315
1870s	750,463
1880s	743,737
1890s	746,595
1900s	757,257
1910s	350,376

Table 3. Old Bailey Corpus: spoken words per decade

[23] An updated release of the OBC (Version 2.0), comprising more than 25 million words, was published in May 2016. At this point, the data extraction for the present study had already been completed.

Only the first and last decades deviate noticeably. The figure is lower in the 1720s because verbatim reporting of speech was only starting to be more common at that time (Huber 2007). *Proceedings* before 1720 contained almost no direct speech, which is why they are not part of the corpus. The 1910s do not reach the target amount of words because the *Proceedings* were discontinued in 1913, which led to a shortage of material for this decade.

The spoken passages in the OBC were semi-automatically identified and annotated with sociobiographical, pragmatic and textual information for each utterance.[24] Specialist software[25] was used to streamline the identification of spoken passages in the proceedings and the extraction of relevant sociobiographical information from the text. However, it was crucial that members of the project team with relevant background knowledge of the social history of Late Modern England engaged in closer reading of the trials, checked the extracted details and added missing annotation.[26]

The OBC provides annotation on several levels: sociobiographical, pragmatic and textual. This makes the OBC the largest diachronic collection of spoken English with this detail of utterance level annotation. Sociobiographical information includes gender, age, occupation and social class. Pragmatic information refers to the role of a speaker in the courtroom, e.g. whether someone was present as a defendant or a lawyer. Textual information comprises information on the scribes, printers and publishers of a given issue of the *Proceedings*. Not all information could be reconstructed for all utterances: out of all the words uttered by trial participants in the entire OBC, about 98% are annotated for speaker gender, 64% for social class and 87% for a speaker's role in the courtroom.

Among the social factors, it was relatively straightforward to assign information on gender and age. Gender was assigned based on traditionally gendered

[24] For more information on how spoken passages were localised and tagged, see Huber (2007) and the section "About the project" on the OBC website (Huber et al. 2012).
[25] The unpublished software, referred to as The Old Bailey Tagger in Huber (2007), was programmed by Magnus Nissel.
[26] The assistants involved in the creation of OBC 1.0 were: Oleg Batt, Carolin Beinroth, Daniela Breitenbach, Michel Eberhardt, Florian Eishold, Eva Kapp, Olga Koslowski, Christina Krämer, Sven Langbein, Patrick Maiwald, Manuela Maus, Veronika Molke, Manuel Müller, Bridgit Fastrich, Sonja Petri, Andreas Reuter, Ulrike Schneider, Nora Schunert, Vikram Singh, Andrea Stütz, Alexandra Tran, Sumithra Velupillai, Janina Werner, Bianca Widlitzki, Julie Wunderlich, Mira Zander-Walz, and Jessica Zesche.

first names and titles like *Mr* and *Mrs*,[27] and age was extracted from the proceedings where mentioned. In contrast, the assignation of occupation and social class merits some further discussion, as there are many possibilities to go about this, usually varying between individual corpora (Kytö 2012: 1522). Annotation on occupation in the OBC follows the Historical International Standard Classification of Occupations (HISCO, van Leeuwen et al. 2002). In this scheme, each occupation title is assigned to a micro group with a name and a number. In the OBC, the micro group's name appears as 'HISCO label' and the number as the 'HISCO code': a speaker owning a pawnshop would bear the label 'Pawnbroker' and the HISCO code '49020'. The code is the basis for the HISCO-derived social class scheme HISCLASS, which integrates the dimension manual vs. non-manual work, skill level, supervision and sector to arrive at a person's classification (van Leeuwen & Maas 2011: 26). The HISCLASS system is available as a 13-level class scheme; a condensed 7-tier version is used for annotation of the OBC (see Table 4).

HISCLASS: 13-class scheme	**HISCLASS: 7-class scheme used in the OBC annotation**
1 Higher managers	1 Higher managers and professionals
2 Higher professionals	
3 Lower managers	2 Lower managers and professionals, clerical and sales personnel
4 Lower professionals, clerical and sales personnel	
5 Lower clerical and sales personnel	
6 Foremen	3 Foremen and skilled workers
7 Medium-skilled workers	
8 Farmers and fishermen	4 Farmers and fishermen
9 Lower-skilled workers	5 Lower-skilled workers
10 Lower-skilled farm workers	6 Lower-skilled farm workers
11 Unskilled workers	7 Unskilled workers, unskilled farm workers and unspecified workers
12 Unskilled farm workers	
13 Unspecified workers	

Table 4. The HISCLASS system

For most quantitative linguistic studies, either scheme is too detailed. However, the system is a good basis to build on.

[27] This is obviously a very simplistic way of looking at gender, driven by technical considerations and the capabilities of automatic annotation. For a more substantial discussion of the variable gender, see 3.4.2.

A speaker's occupation (and thus eventually social class) in the *Proceedings* is recovered either based on their own words (1) or based on information added by the scribe (2) – the latter is only available for the latest trials.

(1) JOSEPH MALKIN sworn. I am a tailor; I live in Kirby-street, Hatton-garden. (OBC, t17810222-28-69)

(2) ANNIE MILLS, barmaid, "Phoenix" public house, Norton Folgate. (OBCProc, t19100111-10)

Two methodological 'shortcuts' were taken with assigning occupation and class: if someone was retired from a particular job, they were treated the same as people working in the job in terms of HISCO and HISCLASS labelling. Where it was not possible to ascertain a woman's job (because she was not asked, did not mention it or because she did not pursue paid work) but information on the husband's job was available, she received the same HISCO and HISCLASS as her husband.

The text in the OBC is based on the digitised versions of the *Proceedings* created for the 'The Proceedings of the Old Bailey, 1674-1913' website. This text differs in some small details from the printed *Proceedings*: for instance, the difference between the long and short *s* in print is neutralised in the digital version. In addition, the formatting and layout of the proceedings is not retained 100% in the digitised versions: while e.g. paragraph breaks and capitalisation are captured, italics and boldface are lost in the digitisation process. For the current purpose, i.e. a study on morphosyntax, these issues are of no consequence. In any case, page images of the printed *Proceedings* are accessible on the project website for 'The Proceedings of the Old Bailey, 1674-1913', so that the original printed text can always be consulted when in doubt about a particular example or concordance line.

3.2.2 *The Proceedings* as a publication in its historical context

The *Proceedings of the Old Bailey* was the title of a periodical published (with a few exceptions) after each sessions (meeting of the court) between 1674 and 1913, which is why they are also known as the *Sessions Papers*. During their 239-year history, the publication had to perform a delicate balancing act to satisfy their two major stakeholders: their paying customers and the City of London authorities. Both had an impact on content, composition, tone and language of the proceedings – sometimes to

a greater, sometimes to a lesser extent. It is therefore important to consider the historical context of the publication and its impact on the textual material.

The *Proceedings* started as a commercial venture in 1674. At the time, crime literature was immensely popular, so printers sent shorthand scribes to court to report on the trials (Emsley et al. 2015k). The early *Proceedings* were not necessarily comprehensive and rather sensationalist and judgemental in tone; entertainment value was the publisher's main concern (Emsley et al. 2015a). Between 1729 and 1778, the publication saw a phase of commercial expansion: the *Proceedings* were expanded to 24 pages, supplemented with advertisements and enriched with more detailed verbatim testimony. These measures were in part meant to fend off competition from alternative compilations of trials and the growing number of daily newspapers (Emsley et al. 2015a). The shift in reporting style in the 1720s was accompanied by a corresponding shift in target audience: judging by the advertisements, the price of the publication and the other publications produced by the various printers of the *Proceedings*, a middle- and upper-class readership was intended (Shoemaker 2008: 563, Ward 2014: 25–26). Comments such as the publisher's remark in 1727 that the *Sessions Paper* was not meant "to please the vulgar part of the town with buffoonery, this not being a paper of entertainment" also suggest a focus on a 'respectable' readership (Shoemaker 2008: 565). Apart from this intended audience, though, the proceedings also reached "lower-class Londoners, including those accused of crimes" (Shoemaker 2008: 575).

As early as 1679, the City of London had got involved in the publication when the Court of Aldermen took the decision that accounts of trials at the Old Bailey could only be published after being ratified by the Lord Mayor and the other justices present (Emsley et al. 2015k). During the 18th century, the authorities "kept an occasional eye on the content" (Shoemaker 2008: 565), but it was not until the end of the century that they meaningfully increased their regulatory role. The 1770s saw a significant increase in City involvement. From 1775 onwards, the *Proceedings* were published under the authority of the Recorder, the chief judge and sentencing officer at the Old Bailey (Shoemaker 2008: 561). In 1778, the City stipulated that the *Sessions Paper* should "contain a true, fair, and perfect narrative of the whole evidence upon the trial of every prisoner, whether he or she shall be convicted or acquitted" (City Lands Committee 1778: 142–143), and not simply on those with entertainment value that would sell well

(Devereaux 1996: 468). These measures were partly a consequence of the Recorder's practice of using the *Proceedings* as a quasi-official record: his recommendations for pardons were based on their reports (Devereaux 1996: 471, Emsley et al. 2015k).

This led to considerably longer and more detailed trial accounts in print (Devereaux 1996: 467) but at the same time slowed down the production process (Devereaux 2007: 19) and increased the publishing costs (Emsley et al. 2015k). The publication was no longer able to fulfil its role as "an up to date purveyor of covertly sensationalist true crime" (Devereaux 2007: 19). In 1787, the City of London had to start subsidising their publication (Emsley et al. 2015k). From the 1780s onwards, the readership of the *Proceedings* consisted mostly of lawyers and officials (Emsley et al. 2015k). After the Criminal Appeal Act in 1907 made the taking of full shorthand notes of trials a statutory requirement, it was only a matter of time before the *Proceedings* died. With an official record existing, there was no need to publish the *Sessions Paper*. In 1913, finally, the publication was discontinued. George Walpole, the publisher at the time, reported that he had only 20 subscribers for the *Proceedings* at the end and had incurred a minus of about £200 a year (Emsley et al. 2015k).

The fact that the *Proceedings* needed to be ratified by official bodies almost throughout their entire history is significant: it is clear that the authorities made use of their power over the material to be included in and excluded from the publication with the aim of portraying a particular picture of crime and criminal justice to the readers.[28] Devereaux (1996: 491–492), for instance, shows that shifting ideologies in the late 18th century led to different reporting practices of trials: while cases ending in acquittals were included in the 1770s to reassure the public of the fairness of the judicial process, the City ordered the exclusion of acquittals from the published record between 1790 and 1792 to reassure citizens that the law would be upheld and transgressors severely punished (Devereaux 1996: 502, Shoemaker 2008: 567). While partly a measure to reduce printing costs, it was also a reflection of the changes in the political climate, heated up e.g. by the start of the French Revolution in 1789, and a shifting view of crime.

The idea of an irredeemable 'criminal class' gained ground throughout the Late Modern period: in the 1780s at the latest, crime was no longer seen as something that

[28] Of course, interference by scribes or editors based on ideological grounds (political or religious) was also found in earlier trial transcripts, e.g. in the Early Modern period (Kytö & Walker 2003: 230).

everyone might resort to given the right circumstances, but as the province of 'the criminal', an emerging "social archetype" rather than an individual (Gatrell 1990: 248). It was the powerless in society, the lower classes, who were saddled with this image: from the late 18th century onwards, their supposed lawlessness was a growing concern to established elites (Gatrell 1990: 243–244). This outlook had been developing for some time: as early as in the 1730s, the printers of the *Proceedings* had started referencing earlier trials to indicate when defendants or their witnesses had previously appeared at the Old Bailey, which contributed to the perception of a "criminal fraternity" that repeatedly landed in front of a judge and needed to be brought to justice (Shoemaker 2008: 574). Increasing social anxieties in the 19th century only intensified the view of a criminal underclass - a view that was firmly established by the 1820s (Gatrell 1990: 251).[29]

The idea that society as such was faced with a worsening 'crime problem' took hold in the 18th century and was also supported by crime reporting in the print media, which grew increasingly serious in tone (Ward 2014: 23) and disproportionately concentrated on the most serious crimes. Ward (2014: 79), for instance, shows that violent thefts were "six times more prominent in press reporting than in the courts" between 1746 and 1750, which thus provided a skewed image of crime to the public. The *Proceedings'* emphasis on testimony in serious cases, often involving violent crime, was likewise used by authorities to emphasise the severity of the apparent crime problem (Shoemaker 2008: 567).

The exclusion of large parts of the case for the defence during parts of the *Proceedings'* history are also to be seen as reactions to the fear of crime and concerns about upholding social order: in 1805, the publisher was ordered to leave out arguments by the defence council;[30] defendants' own statements and character witnesses for the defence were also routinely omitted (Shoemaker 2008: 570). In general, readers of the trials were presented with little evidence of the involvement of lawyers and the argumentative process in court. Shoemaker (2008: 572) considers

[29] A further important shift took place around 1840: crime was no longer simply the "convenient vehicle for the expression of social change" that it had been from the late 18th century onwards (Gatrell 1990: 244), but was actually considered a symptom of change and of a society that was losing its traditional values – a recurring motif to this day (Gatrell 1990: 251–252).

[30] From the late 18th century onwards, defence lawyers appeared in greater numbers (Emsley et al. 2015f). Activity of counsel on the behalf of defendants was however technically illegal until the Prisoners' Counsel Act of 1836 (Devereaux 1996: 500). Before this act passed, defence lawyers were not allowed to address the jury with a summary of the case.

these omissions "ideologically significant" as they perpetuate the idea that justice is unproblematic. The judicial process is portrayed as straightforward and reliable, ending with criminals receiving their just deserts (Shoemaker 2008: 579).

Another area in which the City made use of its regulatory power was when it came to 'indecent' material. Emsley et al. (2015b) report that accounts of trials for sexual offences were restricted during the 18th century: evidence given in trials concerning rape, sodomy[31] and thefts by prostitutes from their clients became less explicit in the course of the century, e.g. by censoring words that were considered obscene (see Widlitzki & Huber 2016 on censoring of taboo and profanity in the OBC). From 1785 onwards, the testimony in sodomy cases was suppressed altogether; testimony in rape cases was no longer printed from 1789 onwards. The following remark on a sodomy trial from 1790 illustrates this point:

> JOSEPH BACON and RICHARD BRIGGS were indicted for committing an unnatural crime. The evidence on this trial, which was utterly unfit for the public eye, did not amount to sufficient proof of the crimes for which the prisoners were indicted, and they were accordingly BOTH ACQUITTED.
> (OBC, 17900224)

The use of the words "unnatural crime" is also a symptom of the cultural norms of the time: sodomy was no longer mentioned by name, but paraphrased as an "unnatural" or "detestable crime" or abbreviated "b-g-y" or "b--y" (Emsley et al. 2015b).

Finally, it has to be said that a lot of spoken testimony never made it into the published *Proceedings* simply due to restrictions of time and space. Shoemaker (2008: 560) shows this by way of an example: the trial of Richard Savage, accused of murder, lasted eight hours but the account in the *Proceedings* is less than 2,500 words long and "could easily have been spoken in under an hour". Huber (2007: 3.2.3) includes a comparison of a trial in the *Proceedings* (trial number 17591024-27) with an alternative account of the same trial published elsewhere. The two accounts show differences in content and language (for a discussion of the latter, see 3.2.3).

The shorthand writers sent to the Old Bailey were aware that not everything was going to make it to print and thus took down only a partial transcript of the

[31] In England, homosexuality was criminalised throughout the period under investigation, as Emsley et al. (2015c) note: "until 1861, all penetrative homosexual acts committed by men were punishable by death". The punishment was reduced to life imprisonment after that date and from 1885 onwards to prison sentences of up to two years. In parts of the United Kingdom, the criminalisation of sex between men continued until 1982.

evidence: they left out or summarised what they considered repetitive or negligible, such as witnesses deposing to the same effect (Devereaux 1996: 480, Shoemaker 2008: 566). Example (3) contains testimony by two persons, but the statement by the second witness is summarised in just one sentence (see boldface).

(3) Crosby. I was collecting the poors rates, in New Inn on the 12th of December, while I was standing at the bottom of No. 1. a person let himself out of a window, he brushed my back, got up and run away, this made me suspect there was something wrong, I went up the stairs, and upon the landing place I found the prisoner sitting upon a sack, we secured him, and upon examining the sack, we found it contained the things mentioned in the indictment.

Mr. Crosby's evidence was confirmed by a Gentleman who was with him.
(OBC, t17750111-4)

What exactly the second witness said is not recorded, and nothing is known about him except that he was "a Gentleman who was with [Mr. Crosby]".

Question-answer sequences to elicit testimony were often summarised as though the witness had produced one coherent text, and cases that were presumably of little entertainment value to readers were condensed to a couple of lines (Shoemaker 2008: 566–567), as in (4):

(4) Susan Fan, of S. Katharine's, was indicted for stealing a Blanket, a sheet, and a Pillow, the Goods of William Shaw, on the 5th of this Instant April. Guilty val[ue] of 10 d. Transportation.
(OBC, t17250407-12)

This is the entirety of the information on Susan Fan's trial, conviction and sentencing to transportation recorded in the *Proceedings*. In light of examples like these, one is inclined to side with Langbein's (2003: 185) assessment that the proceedings were "omitting most of what was said at most of the trials reported". At the very least, it is clear that the *Proceedings* present an incomplete, even selective account of the events in the Old Bailey's courtroom.

However, what was in the reports seems to have been mostly accurate. As Archer (2013: 262) points out, the *Proceedings'* reputation "would have quickly suffered had the reports been largely invented or significantly distorted" – after all, the court was a public place, and the City authorities had a closer eye on the proceedings

from the late 18th century onwards. The coordinators of the 'Old Bailey Online' database also come to this conclusion: their comparisons of the *Proceedings* with other trial reports in manuscript and published form show that "what they [the *Proceedings*] did report was for the most part reported accurately" (Emsley et al. 2015b).

Huber (2007: 3) reminds us that it is necessary to make a distinction between historical and linguistic reliability, the latter of which needs to be measured by different criteria than the former. For the present case, the omission of material as outlined above does not directly influence the analysis of the linguistic variables, although the context of the publication and its history is important to know for the analyst.[32] Of primary concern for any linguistic analysis, however, is whether those parts that are reported are close to what was actually said and the way it was said. This concern will be addressed in the next section.

3.2.3 The OBC as a record of spoken interaction (in court)

The linguistic material in the trial proceedings was shaped both by the courtroom situation as such and by the recording and publishing process. The courtroom situation would have directly influenced speakers' language use, and the production process would have affected the representation of language in print. Both issues are addressed here, and their impact on the usefulness of the *Proceedings* discussed.

Concerning the first point, it is clear that the courtroom setting is a special, rather formalised situation. Everyone has to adhere to a certain protocol in terms of their conduct, including their linguistic conduct. In a trial situation, a person's role in court clearly prescribes when they can speak, and what they can say. For instance, witnesses usually produce past narrative – the OBC thus contains many more past tense verbs than present tense verbs.[33] They very rarely ask questions – that is the domain of lawyers, judges and at times victims prosecuting the alleged perpetrators of crimes themselves, as was customary before the widespread appearance of lawyers. Of course, the courtroom roles also come with a power structure: the dynamics privilege

[32] It is easily visible, though, that other types of studies would run into difficulties: for instance, turn-taking in the courtroom could not be reliably analysed on the basis of proceedings which routinely summarise question-answer sequences.

[33] For lexical verbs, the figures are the following: Lexical verbs tagged VVZ (3rd person singular) and VV0 (base form): 173,139; lexical verbs tagged VVD (past tense form): 862,773.

judges and lawyers. This is even truer for Late Modern than for contemporary courtrooms.

The atmosphere in a Late Modern courtroom was quite different from a contemporary one, especially at the beginning of the period. To name one of the most notable differences, the trials were over much faster than today: sometimes it took less than 10 minutes from the reading of the charges to the verdict. Instead of withdrawing, the jury often deliberated in the courtroom, and jurors made their decisions rather quickly (Emsley et al. 2015f). Late Modern trial procedure disadvantaged defendants (Emsley et al. 2015f): for instance, they were not informed what specific evidence would be presented against them in court and thus had to react spontaneously to witness testimony. The courtroom could also be rather loud and crowded with spectators, which could be a very intimidating experience (Emsley et al. 2015h).

It is well-known that situational factors impact language use.[34] The formality of the trial situation arguably restricts the use of informal language, perhaps especially among those speakers who act in a professional capacity in the courtroom (such as lawyers and judges). However, this does not mean that legal professionals only converse in legalese, i.e. formal and technical language. According to Aronsson et al. (1987), legal professionals tend to switch between styles, depending on who they are interacting with in the courtroom:

> In the dialogical phase (the examination), the legal professionals accommodate to the weaker party (the defendant) by using more colloquial and less formal language. Conversely, our data show how the legal professionals use more legalese (more formal and more technical language, more closely linked to the written documents of law) in the more monological phases of the trial. (Aronsson et al. 1987: 113)

That informal or vernacular features also occur in courtroom interaction is for instance shown by Ching's (2001) case study on a Texan judge's language choices: the judge used the informal and regional plural form *y'all* in the courtroom along with the standard form *you*: *y'all* especially served as a positive politeness device and to establish cordiality. Historical linguists have also pointed out the usefulness of trial data in capturing more oral styles and spoken features. For instance, Kytö &

[34] It is beyond the scope of this study to review the literature on this issue. The reader is referred to Labov (1972b) as a seminal work discussing speech style and to Bell (1984) and Bell (2001) on audience design.

Smitterberg (2006: 209) report that 19th-century trials "displayed features familiar from present-day conversation". At the same time, they were also characterised by language use typical of the courtroom situation (Kytö & Smitterberg 2006: 209).

Apart from the courtroom situation and speakers' consciousness of the situation interfering with language choices, the second factor to consider is the way in which the *Proceedings* were created, i.e. what happened between uttering words in the courtroom and putting them to print. Huber (2010) provides an overview of the production process of an issue of the *Proceedings*, reproduced in Figure 4.

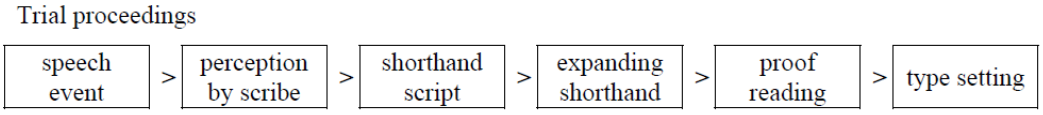


Figure 4. Publication process of *Proceedings*

From the speech events in court to their written record, several intermediate steps are taken: shorthand scribes sent to the trials took notes of the speech events; this shorthand notation was probably expanded to longhand later, proofread, typeset and printed. Based on what we know about the dates the sessions took place and the subsequent publication dates of the proceedings, we know that the entire process cannot have taken more than a couple of weeks (Huber 2007: 3). Archer (2013: 262) points out that the speedy publishing process of the proceedings automatically ensures that they comply with one of the recommendations formulated in Kytö & Walker (2003: 241–242) that are supposed to help limit editorial interference with the linguistic material, namely to prioritise early imprints of trial records over later versions or editions. Apart from the fact that transformation into the written mode necessarily cannot retain all aspects of the spoken mode (see Section 2.2.2), each of these steps also represents a possibility for interference with the linguistic material.

In the first step of the process, scribes observed the court proceedings and took notes. The scribes worked under rather difficult conditions, as indicated above: most trials were over fast, and the courtroom could be quite noisy. For the *Proceedings*, it can be assumed that the scribes used shorthand notation: from 1749 onwards, we can be sure of it, but it is likely that the practice started with the first proceedings in the 1670s (Huber 2007: 3). After all, shorthand was already used in earlier trials, as

Culpeper & Kytö (2010: 51) point out. The scribes' perception of speech events represents the basis of the OBC. There is of course, a possibility that a scribe misheard certain things, considered some aspects unimportant or treated some linguistic forms as interchangeable. Huber's (2007: 5) comparison of a trial as it appeared in the *Proceedings* with an alternative account of the same trial published elsewhere finds "some verbal overlap" but also "substantial differences, including omissions as well as verbal, morphological, and syntactic divergences". It is impossible to know which version of the trial resembles the utterances in the courtroom most accurately. While this does not constitute an argument against the linguistic reliability of the *Proceedings* as such, it plainly illustrates that they should not be read as 1:1 representations of the spoken word.

Two statements by Thomas Gurney, a scribe who worked on the *Proceedings* in the mid-18th century, on his note-taking practices are instructive here. In 1754, Gurney was sworn as a witness in the perjury trial of Elizabeth Canning. After reporting on Ms Canning's evidence, he was asked whether he was "able to say, upon [his] oath, that that was the evidence that the girl, upon her oath, then gave in court", to which he replied "The substance of it is the evidence she gave in court" (Howell 1816: 326, see also Shoemaker 2008: 566). That there was apparently a difference between the substance of the evidence represented in the shorthand notes and the exact utterance in court is also clear from Gurney's treatment of nonstandard English. When Moses Henericus was tried for perjury in 1758, Gurney was once again called as a witness to recount the defendant's statement in a prior trial. Gurney presented Mr Henericus' evidence in standard English, and added: "'I took that to be his meaning which I have printed, he speaking as most of the foreign Jews do, a sort of broken English" (OBCProc, t17580113-30; see Huber 2007: 3).

Nonstandard pronunciation or morphosyntax were not necessarily transformed into standard orthography by the scribes, though. Sometimes, there are efforts to retain the original morphosyntactic choices and also to visually represent the pronunciation of speakers of dialects or speakers using English as a foreign language (Huber 2007: 3). A number of such instances exist in the earlier proceedings: 63 passages could be

identified in the OBC between 1720 and 1757.[35] They include foreign nationals like Julian Brown (5), who identifies himself as an Italian citizen and French speaker when asked, and speakers of English with a notably Welsh (6) or Scottish (7) accent.

(5) [Julian Brown] I no can tell vat Day - But ven I come dare I meet Mr. Ruffhead, and he vant Mr. Campbell too. So ve go togader to de Stag and de Hound, and dare I find him and de Presonaar - De Presonaar vas in de Vite Grey Coat vid de Button Silver and de Scarlet Vaistcoat vid de Lace upon it. (OBC, t17350702-22)

(6) [Lewis Jones] Then the Prisoner struck the Teceaset with the flat Site of his Packonet, and sait, stand by; and then the Teceaset took his Fist a thuss'n, and went to the Prisoner - I think you call it Tarting, for I call it so in my Welch Way; ant then the Prisoner took his Swort so - and stapt him in the Left-side [...]. (OBC, t17311208-48)

(7) [Robert Johnston] What is the Matter with you now, Sir? says Mr. Taylor Eloard, Sir, I have been roab'd of my Sword, says I; and wha has taken it from ye, sir, says he; why, that Fallow, Sir, says I, that pratanded to get my Box for me, kenye what his Name is? O give me a Pan and Ink, and I'll sat ye doon his Name, and what he may find him. (OBC, t17320525-30).

It is true that (6) seems to be exaggerated for comic effect (for another example, see Huber 2007: 3), especially as this witness is accorded much space in the trial account (669 words) and is made to look a bit bumbling in front of the court in general. The line "for I call it so in my Welch Way" was likely added by the scribe. Examples like (7), which includes Scottish English features like e.g. the lexical choice of *ken* (*ken ye what his name is?*) or the visual representation of the Scottish English pronunciation of *down* (<doon>), are more neutral and part of a serious questioning of the witness without any comedic overtones for the benefit of the readers. Such representations of nonstandard usage may be considered an indication of linguistic faithfulness (Huber 2007: 3) and show that the scribes were "aware of the linguistic practices of participants in the trials, and thought them worthy of representation" – even if this representation was not always entirely accurate (Traugott 2011: 72).

[35] During the annotation procedure, some passages were identified as notably nonstandard English by the assistants. These were tagged <nonstandard> in the OBC and can be searched for in the offline version of the corpus. As they came to the project team's attention during close reading of the trials, the tagged instances do not represent an exhaustive list of all nonstandard passages in the texts.

There are further helpful indicators for assessing the linguistic reliability of speech-related texts such as the *Proceedings*. When it comes to trial proceedings and witness depositions, Kytö & Walker (2003: 241) conclude that texts that are "fairly reliable records of spoken interaction" are characterised by references to the use of shorthand, the reporting of slips of the tongue, dialect words, and the representation of pauses and interruptions. Having already established that the *Proceedings* made use of shorthand and that dialects were partly represented, let us turn to slips of tongue, pauses and interruptions. These do appear in the published accounts, but not consistently. Fillers like *uhm*, *ah*, are not reported, but interruptions and hesitations are visible, as shown in (8) and (9).

(8) [Elizabeth] Berry. **After I got off the Bed –**
Q. Off the Bed!
Berry. Off the Ground [...].
(t17431012-15)

(9) The Prosecutor [William Hopkins] depos'd, that it being Sunday Night, at 10 or 11 o'Clock, he went into a Brandy-Shop in Drury-Lane, and finding the Prisoner there, he treated her with a Quartern of Geneva, that inspir'd by this he went with her into a Back-Room, **and there – and there – and there –** she pick'd his Pocket; [...]. (t17280228-41)

In (8), Elizabeth Berry's contradictory statements prompted a lawyer to interrupt her mid-sentence, shown by use of the punctuation (dash). In example (9), the hesitations in the (indirectly reported) evidence by William Hopkins are probably due to the witness trying to find a delicate way of expressing that he went off with a prostitute who most likely stole from him. His hesitations are on record – even though his testimony is not recorded as direct speech. Slips of the tongue are also sometimes reported, including scribal/editorial corrections in brackets.

(10) [Elizabeth] Cresswell. I could not **preserve (observe)** him by Day Light, but there was a Light at my Master's Door - And I saw him next Morning in the Compter in the same dress.
(OBC, t17350522-32)

(11) [Philip Price, watchman] I know nothing of the Robbery, but as Mr. Brown had **subscrib'd (describ'd)** the Man, I call'd the Constable, and we went and took the Prisoner, as he had **subscrib'd** him, in Thatch'd Alley, in Chick-Lane.
(OBC, t17400116-4)

Note that in (11), the second time the speaker uses the wrong lexical item (*subscribe* instead of *describe*), it is not corrected. Occasionally, as in (12), we also find corrections of syntactic choices.

(12) I can't say **nothing** (any thing) in his Behalf [...].
(OBC, t17401015-57)

In this particular case, the target of correction (and perhaps implied criticism) concerns the form of the negation. In general, the practice of adding corrections in brackets is advantageous for linguistic research because the original utterance is preserved. It is striking, however, that all these examples of dialect data, hesitations and corrections are found before 1760. It is possible that with the increasing control of the City of London, which took a significant step in 1775, editorial interference grew.

Support for the impression of growing standardisation comes from Huber (2007) and Huber (2010), two studies on contraction. Contractions are assumed to be diagnostic of spoken language. Huber (2007) uses the rate of contraction in the *Proceedings* as one way of assessing internal consistency, i.e. whether a variable feature is consistently portrayed in a similar fashion in a given corpus, and external fit (see Section 2.2.2; see also Schneider 2002, 2013). Micro studies of negative contraction in short subperiods, in which either the scribe or the printer varied (1751-1761, 1780-1782, and 1795-1805), revealed differences in the treatment of negative contraction by different scribes and printers, either across the board or with regard to individual auxiliaries (Huber 2007: 5). However, a follow-up study clarifies that "the overall development of negative contraction in the OBC is too regular to be attributable to the influence of [the scribes, printers and publishers of the *Proceedings*]" (Huber 2010: 72). Instead, variable negative contraction in the *Proceedings* is best explained as the result of efforts to "maintain maximal phonological contrast between positive and negative polarity" (Huber 2010: 78), which is a point in favour of the internal consistency of the OBC. However, the overall contraction rate in the OBC is much lower than expected for supposedly spoken language and fails to show the expected pattern of increase for an incoming feature.[36] Instead, a relatively high rate in 1732-1759 (16.9%) is followed by a fall to 4.0% in

[36] It was in the second half of the 17th century, i.e. just before the time span represented in the OBC, that cliticisation of the negator *not* onto its auxiliary host was established in writing (Mazzon 2004: 104–105, see also Huber 2010: 68).

1760-1789 (Huber 2010: 68). Huber (2010: 69) argues that external pressure in the form of growing City control and the shift to increasingly formal style is responsible (see Section 2.2.3 for 18th century opinions on contractions), and states that "the rate of negative contraction was much higher in the language used in the Old Bailey courtroom than documented in the *Proceedings* from 1760 onwards". While this makes it obvious that the OBC does not reflect speech completely faithfully, a comparison of OBC trial accounts with trials in the CED in the period of overlap of these two corpora (1720-1760) shows that the rate of negative contraction is similar. Huber (2007: 3) concludes that there is a high degree of external fit and that "[t]he OBC can therefore be taken to be just as representative of spoken language as other trial texts".

The next hurdle in the *Proceedings'* production process was getting the shorthand notes into printed form. Although it is theoretically possible to do the typesetting straight from the notes, it is more likely that they were first put in longhand and then proofread (Huber 2007: 3). As no manuscripts of any issue of the *Proceedings* have been discovered (Huber 2007: 3, citing p.c. of 2007 with Tim Hitchcock), it is not possible to check what shorthand systems were used. However, the system developed by Thomas Gurney, who took down the *Proceedings* in the mid-18th century,[37] was very influential. Gurney's system can be assumed to have been in use during his term and that of his son Joseph, who followed him as scribe until 1782. Shorthand continued to be used afterwards.

The step of transforming shorthand into longhand once again presents a potential for errors. Additionally, it requires the scribe to make choices about how to represent some features because the shorthand systems are not designed to capture all linguistic detail. Some examples of linguistic details that Gurney's system was unable to cover are pointed out in Huber (2007: 3): the symbol for the letter *t* could also represent the pronoun *it*, making it impossible to distinguish e.g. between *it will* and *'twill* based on shorthand notes only. The system also makes no distinction between the forms of the indefinite article *a* and *an*. Despite these issues, the fact that shorthand was used at all is a plus for the accuracy of the *Proceedings* in two ways. First, shorthand enabled scribes to record speech practically simultaneously to its utterance,

[37] Gurney was appointed official shorthand writer of the *Proceedings* in 1749, but very likely had recorded trials in an unofficial capacity for some time at that point. He had already moved to London in 1737 (Canadine 2016).

which typically increases the accuracy of written records of speech (Huber 2007: 5, Schneider 2013: 71–72). Secondly, the fact that Gurney used the *Proceedings* as advertisement for his shorthand system was a strong motive to produce accurate transcriptions. (Shoemaker 2008: 563).

After the longhand version was finished and proofread, the *Proceedings* would have been printed. As the City authorities had an eye on the published content, it is fair to ask whether the linguistic form was also their concern. This is not directly spelled out in any regulations on the *Proceedings*, but it seems that the City was primarily interested in ensuring the decency and respectability of the *Proceedings*. This arguably pertains mostly to questions of content and not of language.

However, there were times when concerns about language and concerns about content did intersect. In 1725, the printer and shorthand scribe of the *Proceedings* of 7 April 1725 were forced to apologise before the Court of Aldermen for what the Court called "the lewd and indecent manner of printing the last sessions paper" (Shoemaker 2008: 564, quoting Court of Aldermen 1725: 368, 376–377). Huber (2007: 3), focussing mainly on the transcript of James Fitzgerald's evidence in this issue of the *Proceedings* (shown in part in (13)), considers this reprimand an "indication of the control that the City exerted not only on what was reported but also on the language in which it was reported".

(13) On the 25th of February last, about 11 at Night, O' my Shoul, I wash got pretty drunk, and wash going very shoberly along the Old-Baily, and there I met the Preeshoner upon the Bar, […]. Sho we went together; but not having any Deshign to be consherned with her, I paid her Landlady a Shilling for a Bed. For it ish my Way to make Love upon a Woman in the Street, and go home with her, whenshoever I intend to lie alone. […] she wash after being concerned with my Breeches, and got away my Watch […] for fear she should get it from me, I let go my Hold, and went for a Constable, and he carried her to the Watch House, where he took the Watch upon her. He found it in a Plaushe that my Modesty won't suffer me to name; for ash I am a living Chreestian, she had put into her ***. (OBC, t17250407-66)

I personally think that content rather than language was primarily at issue in this case. Certainly, in terms of language, it could have been objectionable that testimony by an Irishman was rendered phonetically for comic effect. It also seems that the Irishman

was set up to look ridiculous in the printed account by putting obvious contradictions into his mouth (e.g. *was got pretty drunk – was going very s[h]oberly*). Much more important, however, would have been that fact that the trial account involved 'indecent' subject matter such as discussions of sex and (probably) prostitution.

The treatment of taboo language in the *Proceedings* also fits this reasoning: it declines in frequency throughout the Late Modern period, and swearwords are increasingly censored in print. Most often, this is done by replacing part of the offensive word with dashes, e.g. by writing *d—n* instead of *damn* (Widlitzki & Huber 2016). It is likely that the authorities kept an eye on such things. However, this censoring strategy leaves enough room for guessing the original meaning and even the original words, so it is not particularly strict in terms of the language.

In the end, it is clear that at OBC texts cannot be considered verbatim transcripts of the courtroom interactions. As Kytö & Walker (2003: 241) put it, "every written speech-related text has been produced by the scribe rather than the speaker(s)". Additionally, the courtroom also impacts speakers' choices. Tieken-Boon van Ostade (2000: 446) speaks of a "double observer's paradox", created by the speakers' register consciousness, in this case regarding appropriate linguistic form in a courtroom, and the subsequent recording of speech in writing. This does not mean that linguistics cannot gain any insights from court records. It is, however, essential to keep the above-mentioned issues in mind in the subsequent analysis.

3.2.4 The OBC as testimony of and to a group of speakers

It is obvious that the Old Bailey Corpus can only offer a small glimpse of the goings-on in the courtroom and into the realities of Late Modern England. Next to the constraints originating in the creation of trial texts (both technical and ideological), it was the justice system itself that had the most influence on the form and content of the texts, as it "determined which trials reached court, which witnesses gave evidence, and whose voice would be heard" (Emsley et al. 2015i).

As the central criminal court for the City of London and the County of Middlesex, the Old Bailey was the site of all trials for serious crimes, i.e. felonies and the most serious misdemeanours, in these areas (Emsley et al. 2015d). With the Central Criminal Court Act of 1834, its jurisdiction was enlarged to also encompass

metropolitan Essex, Kent and Surrey (May 2003: 147). This means that we know where the majority of speakers in the *Proceedings* lived. How much and what kind of linguistic material was produced by participants in Old Bailey trials is mainly dependent on two factors: their representation in the courtroom (how likely were they to be in a courtroom?) and their role in the trial (were they witnesses, defendants, judges, etc.?). Not all persons were equally likely to appear in front of a court and, if they did, to fill certain roles. In fact, factors like gender and class played a major role.

The speakers in the OBC all fill one of the following roles (of which not all occur in all trials): judge, lawyer, witness, victim, defendant, and interpreter. The amount of talk that a person contributes to the *Proceedings* and the kinds of interactions they had with others in the courtroom is obviously partly dependent on their role. Figure 5 shows the word count in the OBC by role.

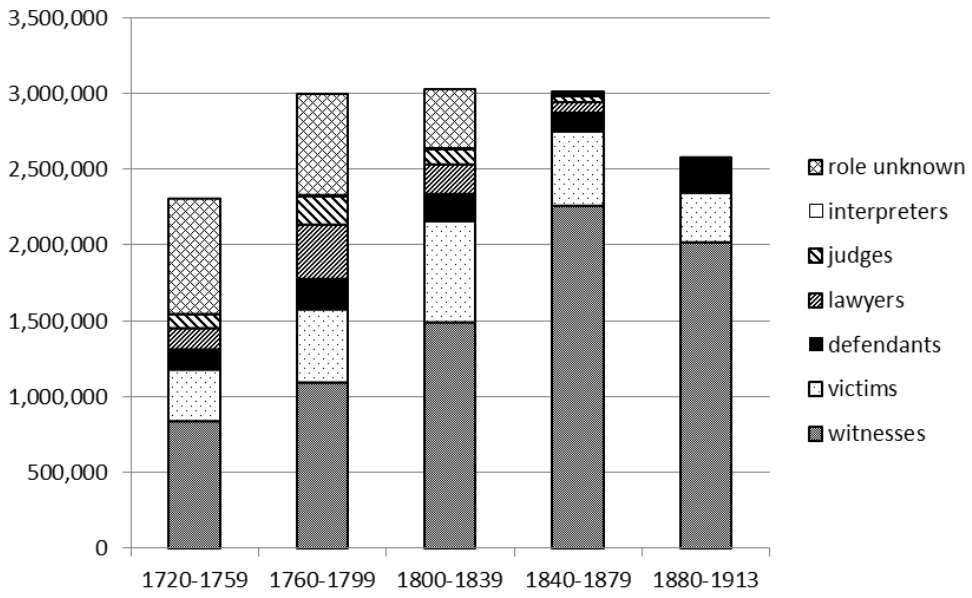


Figure 5. Words spoken in the OBC by speaker role

In all periods, witnesses provide the bulk of the recorded discourse: they produce ca. 55% of all words in the OBC. Since spoken testimony was regarded as the key to establishing guilt or innocence throughout the Late Modern period, and since the readership of the *Proceedings* was presumably most interested in the retelling of the crimes, this is expected. The role of a victim was very similar to that of a witness in the later *Proceedings*, but at the beginning of the period, victims would also act as prosecutors and ask questions of witnesses, etc. – jobs later taken over by lawyers.

That the words of defendants are only partially reported in the *Proceedings* has already been established above, so the low proportions are not surprising. Interpreters have a marginal role: where trial participants are unable to communicate in English, their statements are translated; only the English translation is set down in the *Proceedings*. Throughout the period, judges had supreme authority in the courtroom, but their role shifted from those conducting the trials and leading the interrogation of witnesses in the 18th century to a more backgrounded role of managing the adversarial contest of the lawyers (Emsley et al. 2015f). Lawyers, who only became a regular feature of courtroom interaction towards the later 18th century, also make up for only a small part of the recorded discourse. This makes sense considering that their questions were routinely omitted in favour of summarising witness testimony as coherent texts, and restrictions on providing and reporting on defence council were in place for parts of the period under investigation (see 3.2.2 in general and fn 30).

Gender is another major factor to consider, and is also heavily intertwined with role. In fact, the roles of judge and lawyer are off limits to women. This is part of the reason why only 2.3 million words in the OBC (i.e. roughly 17% of all words where information on speaker gender is available) are uttered by women. This proportion remains relatively stable throughout the decades, as Figure 6 shows.

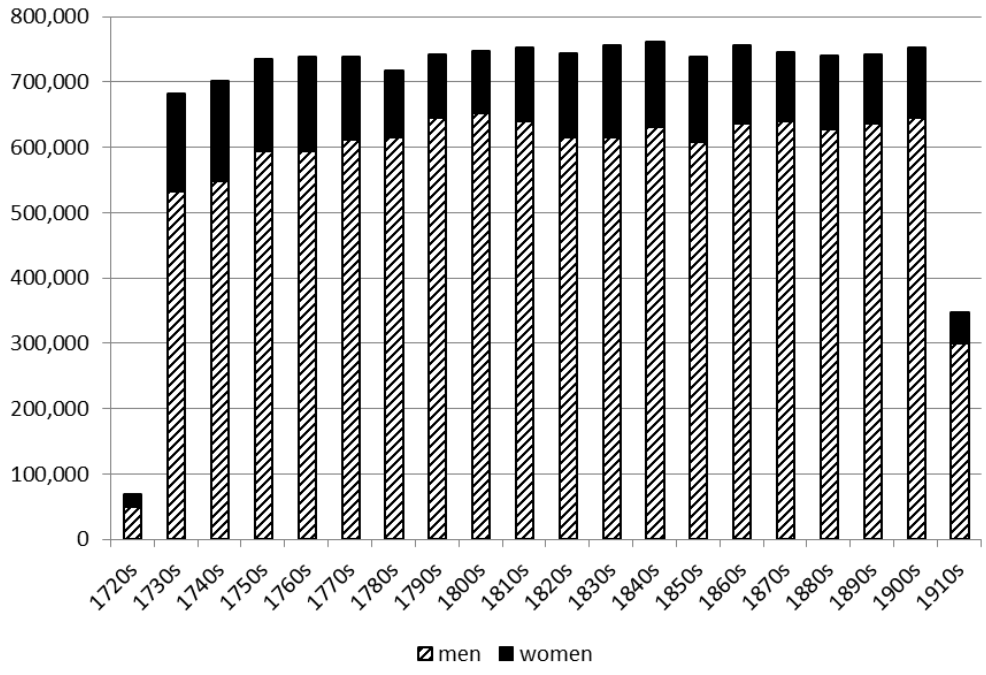


Figure 6. Words produced by gender and decade in the OBC

The only roles remaining for women were those of defendant, victim and witness. Even in these roles, though, the corpus has fewer female speakers and fewer words uttered by women.

The underrepresentation of women in particular roles has a number of interrelated reasons. Out of all the defendants in the entire *Proceedings* from 1674 to 1913, women account for only 21% (Emsley et al. 2015g).[38] Emsley et al. (2015g) suggest that this is a consequence of the social norms and of prevalent ideas about gender and gendered behaviour at the time. In the first place, it is possible that women simply had less possibility to commit crimes that would land them in the dock at the Old Bailey: they usually did not carry weapons or heavy tools and their role was more closely confined to the home than men's. Perhaps more importantly, unsocial behaviour was interpreted differently depending on the gender of the perpetrator. In women, it was considered highly unusual and was therefore less likely to be interpreted as criminal (Emsley et al. 2015g). Crime – along with power and aggressiveness – was considered the province of men. As a consequence, women were less likely to be formally prosecuted. Often, other correctional measures were employed, including informal arbitration, or the cases were prosecuted in minor courts, such as the Quarter Sessions courts. However, where women significantly deviated from their assigned gender role, e.g. by hurting children, they were usually formally prosecuted and severely punished (Emsley et al. 2015g).

Among the victims in all trials recorded in the proceedings, only 10% are listed as women, ca. 60% as men and the rest as indeterminate, unknown or mixed (Emsley et al. 2015j). The explanation for this is mostly a legal one:

> […] theft was the most common offence prosecuted, and most marital property was deemed to be in the possession of the husband. Thus, even if a woman's clothes were stolen, if she was married her husband would have been labelled as the victim of the crime. (Emsley et al. 2015g)

From a legal point of view, then, women were rarely in a position to be victims of theft. Figure 7, which compares the genders of victims across different types of crime, also makes that clear (Emsley et al. 2015j): women make up only a small portion of

[38] This figure is subject to chronological fluctuations: while 40% of the defendants were women between the 1690s and 1740s, this proportion had shrunk to only 22% by the early 19th century and further to 9% by the early 20th century.

theft victims (ca. 10%) and violent theft victims (<15%). They are represented to a greater extent in other categories, like killing[39] (33.4%) or sexual offences (83.8%).

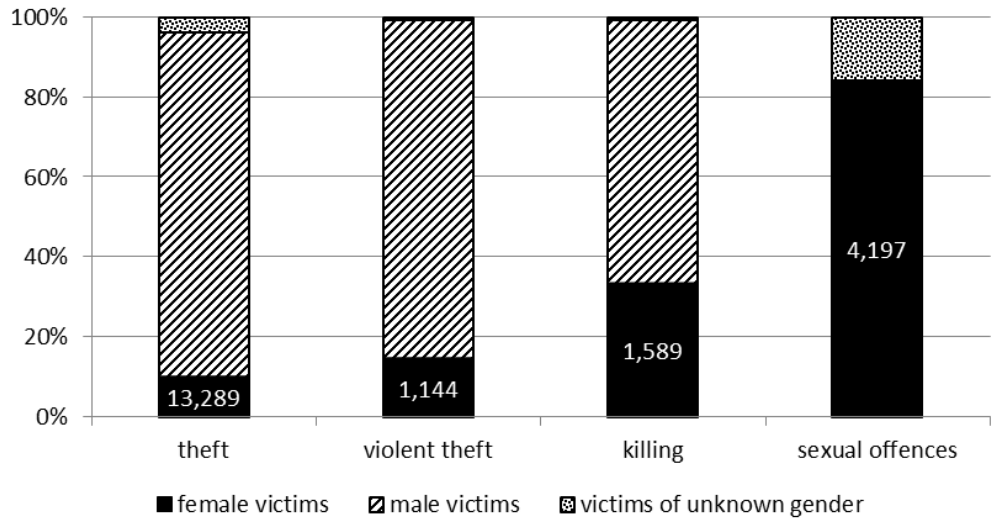


Figure 7. Victims of various crimes by gender in the *Proceedings of the Old Bailey*

In addition, it is possible that the judicial process at the male-dominated Old Bailey deterred single women, who had been victims of crime, from coming forward and prosecuting a case, as Emsley et al. (2015g) suggest. Instead, there is evidence that women were more likely to use "less formal legal procedures such as summary jurisdiction and informal arbitration" (Emsley et al. 2015g).

Another reason for the lower word count produced by women has to do with the kind of evidence women were called to give, and the *Proceedings'* way of dealing with women's testimony. When a woman was a defendant at the Old Bailey, the experience must have been "significantly more intimidating […] than it was for men" (Emsley et al. 2015g), and might have made them reluctant to speak much. When women were called as witnesses or came to prosecute a crime, they were faced with a greater level of doubt: there are indications that "juries treated evidence presented by female witnesses more sceptically than that delivered by men" (Emsley et al. 2015g), which may also have discouraged them from speaking at length. It is also important to know that women's testimony "was more likely to be omitted from the Proceedings" (Emsley et al. 2015g). One reason was that women were often called as so-called

[39] Needless to say, people who were killed do not talk in court, which does not increase the word count for either women or men.

'character witnesses', whose evidence was routinely left out of the records, or summarised as in (14).

(14) The prisoner's aunt gave him a good character.
(OBCProc, t18260511-38)

Evidence that women and men were treated differently when in the witness stand is also found in earlier trials. Culpeper & Kytö (2000: 81) find that women witnesses in Early Modern trials are "constructed by the male judges (and possibly the male scribes) more as crime-narrative givers than are the male witnesses, who are more involved in intense cross-examination". They suggest that women mostly cooperated in filling this conversational role for their interlocutors because of their relatively powerless state (Culpeper & Kytö 2000: 81).

Like coverage across genders, coverage across social strata is not equally distributed. In sum, the OBC over-represents higher-class speakers, as Figure 8 shows.

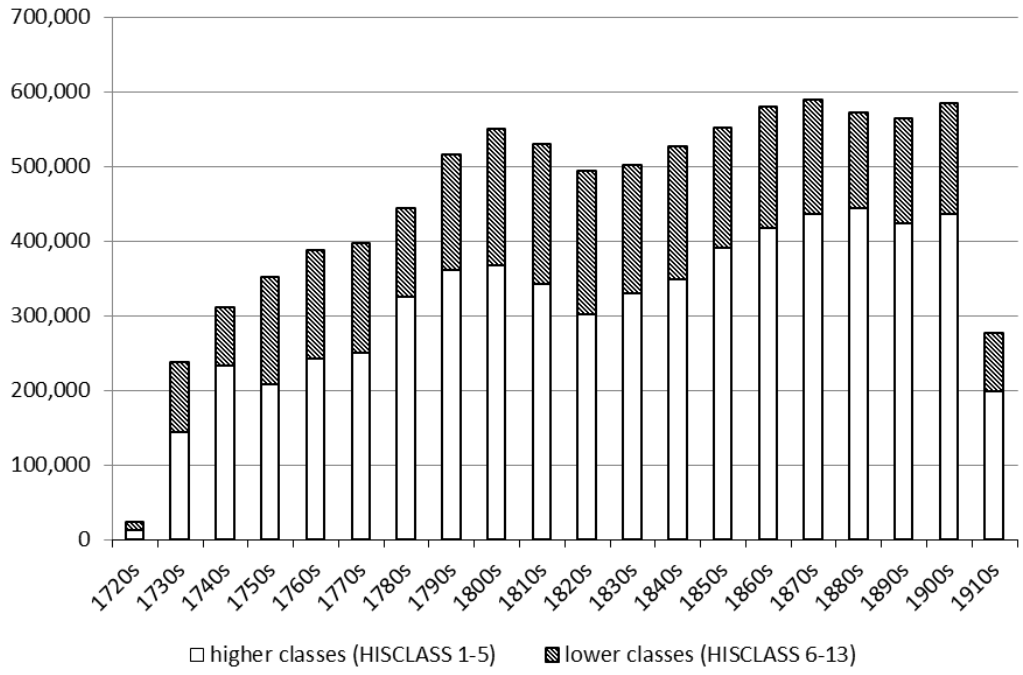


Figure 8. Words by class, OBC

Throughout the period, higher-class speakers (with white-collar occupations, sorted into HISCLASS 1-5 in the 13-tier scheme) are much better represented than those in the lower classes (performing blue-collar occupations, sorted into HISCLASS 6-13).

Ca. 2.8 million words are uttered by lower-class speakers, and roughly 6.2 million words by higher-class speakers.[40] This imbalance concerning social ranks is partly a consequence of certain roles in the courtroom only being available to higher-class men. All lawyers (HISCO: 12110) and judges (HISCO: 12210) belong to the second-highest class in the 13-tier HISCLASS scheme by merit of their occupation.[41] For many other trial participants, information on their occupations is simply not retrievable, which is why 4.9 million words' worth of speech have no annotation for social class.

All speakers mentioned by name in the *Proceedings* have a unique speaker identifier in the OBC, through which all their utterances are recoverable: speaker 17320223_0027 is speaker 27 in the *Proceedings* published on 23 February 1732, for example. An exception is found with certain judges, lawyers, jurymen or interpreters who could not be identified by name. These receive generic speaker identifiers throughout an issue of the *Proceedings*: the identifier 17320223_?CRT, for instance, is assigned to all utterances made by (a) speaker(s) referred to as "Court" in the text of the *Proceedings* of 23 February 1732. It is not possible to say in such cases whether the same judge presided over every trial in a session, so the speaker ID may well cover several persons. Recovering the identity of lawyers is also problematic, as they are usually referred to only as "Counsel" in the text. Jurors are assigned generic IDs, too: while their names are listed in the front matter of each issue of the *Proceedings*, their utterances are prefaced by "a juryman" or "jury", which makes identification impossible. The majority of speakers, however, are clearly identifiable individuals.[42]

Despite the restrictions discussed, the OBC offers a fascinating opportunity to investigate (the written representation of) Late Modern speech. Although the setting is an official one, it affords a glimpse into people's everyday realities. Through the work of the court scribes, we are in the lucky position to gain access to the English of speakers who would never have been able to leave a written record of their language use behind. This gift needs to be handled respectfully. Some people immortalised in the *Proceedings'* pages undoubtedly (re-)experienced some of the worst days of their lives in this courtroom. Despite the official character of the records and their overall

[40] Combinations of underrepresented social factors may compound tendencies observed in isolation: for instance, lower-class women only provide 370,000 words in the OBC.
[41] Jurymen must be assumed to belong to the middle or higher ranks for most of the period (Emsley et al. 2015e). As no detailed information on their background is available, their social class is treated as unknown in the OBC.
[42] When combining generic speaker IDs with the trial identifier (see 1.4), it is possible to further group the utterances by a generic speaker in one issue of the proceedings by trial.

impassive tone when stating crimes ("rape", "murder") as well as punishments ("transportation", "death"), it is clear that what we as scholars study in terms of linguistic substance had an undeniable effect in the real world for the people involved. The nature of the language that we investigate may not be as private as diaries and personal letters in this instance, but demands respectful handling in the way that more personal documents do (see e.g. Dossena 2012: 37–38, Elspaß 2012: 165–166 for a comment on dealing with personal documents in linguistic inquiry).

3.3 Supplemental data: the Corpus of Late Modern English Texts

The Corpus of Late Modern English Texts, Version 3.0 (Diller et al. 2011), short CLMET3.0 (or just CLMET for the purposes of this work), is a multi-genre corpus comprising ca. 34 million words of running text. It is a collection of published works by British authors in the Late Modern period. The texts are all in the public domain and were retrieved from various online archives. The design of the CLMET, which is available with and without POS tags, is outlined in detail in Table 5.

Genre	**1710-1780**	**1780-1850**	**1850-1920**
Narrative fiction	4,642,670	4,830,718	6,311,301
Narrative non-fiction	1,863,855	1,940,245	958,410
Drama	407,885	347,493	607,401
Letters	1,016,745	714,343	479,724
Treatise	1,114,521	1,692,992	1,782,124
Other	1,434,755	1,759,796	2,481,247

Table 5. Design of the CLMET (Diller et al. 2011)

The corpus covers the period 1710-1920, and is divided into three 70-year subperiods. It includes the genres 'narrative fiction', 'narrative non-fiction', 'drama', 'letters', 'treatise' and 'other'.

Multi-genre corpora like the CLMET are usually employed "for diagnostic purposes": they can point out general trends that can then be compared with trends based on data from smaller, more specialized corpora, for instance" (Kytö 2011: 424). In the other direction, they can be used to contextualise results from specialised corpora by allowing a broader base of comparison and e.g. answer the question whether only one genre is behaving in a certain way. In light of what Culpeper & Kytö

(2010: 3) say about the value of triangulation to investigate historical (spoken) language use (also see 2.2.2), the CLMET plays an important role in this study: the results from the OBC will, so to speak, be triangulated with findings from existing research and also with results from a speech-related CLMET subcorpus. For most features, the drama subcorpus (without stage directions) will be used. In one case (Chapter 6), the drama corpus failed to produce sufficient data for useful comparison to the OBC results, so the narrative fiction subsection of the CLMET was used.

3.4 Choosing and defining variables

The present variationist study is interested in establishing whether there are correlations between social factors and the use of particular linguistic variants in the Late Modern period. It aims to put forward explanations of observed variation with regard to the social context of language use. The present section introduces first the linguistic variables under investigation (3.4.1) and then the social variables, their contextual relevance and operationalisation (3.4.2).

3.4.1 Linguistic variables

As mentioned in Chapter 1, four linguistic variables will be considered in the present analysis. They are shown in Table 6 (reproduced from Table 1 for convenience), and will be discussed individually in Chapters 4-7.

linguistic domain	**feature**	**(major) variants**
modality	1) verbal expressions of obligation and necessity	HAVE TO: *You **have to** go.* MUST: *You **must** go.*
auxiliation	2) choice of perfect auxiliary	BE + past participle: *I **am come** home.* HAVE + past participle: *I **have come** home.*
tense	3) choice of narrative tense with SAY	historic present: *so **I says** to her* past tense: *so **I said** to her*
agreement	4) verb form with 2SG pronoun *you*	*was: **you was** alone* *were: **you were** alone*

Table 6. Features under investigation

What all features have in common is the fact that they are grammatical variables, are part of the verb phrase and have been shown to exhibit variation (some involving

change) in the Late Modern period. Importantly, none of these features have been the focus of a detailed corpus-based analysis from a sociolinguistic perspective so far. The choice to restrict the analysis to a small number of variables is deliberate: it allows a greater level of detail, in this case an exploration of sociobiographical, situational and textual variables.

To have some diversity within the group, care was taken to include different linguistic domains and variables of different 'sizes', so to speak. The variables span modality, agreement, tense and auxiliation. In two cases, the variants represent different inflected forms of the same verb (*was-were*, *says-said*). One variable represents a choice between two auxiliaries (BE – HAVE), and in one case, we are even dealing with the choice between two different categories of verbs (modal – modal auxiliary). The variants of the variables also differ in a further respect, i.e. how they were commented on in contemporary grammars. Some received quite a lot of comment (either negative or positive, and sometimes changing), while others hardly provoked the attention of codifiers (details are provided in the coming chapters).

In addition, the variables differ with regard to whether the variation between alternatives represents a change in progress. This criterion divides them into two groups: two alternations represent systematic changes in progress (choice of perfect auxiliary: BE – HAVE; expressions of obligation and necessity: MUST – HAVE TO), while the other two (variation between *was* and *were* with 2nd person singular pronouns and variation between historic present and past tense) do not. The latter two do not develop towards any of the two forms to the gradual exclusion of the other. Instead, the variants *you was* and *I says* are considered features of conversational grammar,[43] or informal features, and are also frequently found in spoken registers today. They rarely occur in other settings, which should make it interesting to see how they do in the OBC, a record of speech. The OBC's speech-related nature is also of interest to the study of the two changes in the Late Modern period: as grammatical change is more frequently initiated in spoken language (see e.g. Leech 2003: 226), it will be interesting to analyse how far along the trajectory of the change is in the OBC compared to earlier studies.

Finally, the variables in question also had to hold up to some practical considerations. At the most basic level, there needs to be variability of the

[43] For contemporary English, Quaglio & Biber (2006: 702, 713) point out that the use of 'vernacular' features, including both widely used colloquialisms and stigmatised forms, is a characteristic of conversational language.

phenomenon analysed in the given textual source for a variationist study to work (Schneider 2013: 60). When pursuing a quantitatively informed study, it is further necessary that "reasonably large token frequencies of individual variants" are present (Schneider 2013: 60). In light of the aim to integrate several extralinguistic variables into the present analysis, this is an important requirement. Keeping in mind the production process of the OBC, it was further important to select only variables whose variants can be distinguished based on the shorthand notation used to first record the speech events (see Section 3.2.3).

When it comes to operationalising the variables in terms of discrete variants, two factors are notable: first of all, Table 6 above shows that all variants are very strictly delimited. For instance, only the verb SAY is considered when it comes to the alternation between historic present and past tense. Such measures are taken in order to keep the amount of data manageable. Secondly, all variables happen to be binary. I purposefully use the phrasing 'happen to be' because having binary variables was not a prerequisite of my research design. I also acknowledge that there are of course further options in the categories, which is why Table 6 lists those variants that are discussed in detail here as '(major) variants'.

Of course, the linguistic variables from Table 6 are to some degree idealisations. In fact, the very idea that we can retrieve all different variants of grammatical variables is an idealisation as these are not a naturally finite set. The process of defining variables and setting up variants is always informed by the choices of the researchers. I have made my choices based on previous work, which I am going to comment on in the respective sections, and based on exploration of the textual material at hand. For instance, I opted to limit the verbal expressions of obligation and necessity to HAVE TO and MUST because these two variants represent the vast majority of strong obligation markers in the OBC. (HAVE) GOT TO, another strong obligation marker, is almost absent: the OBC only contains 43 instances with unambiguous obligation reading (42 of HAVE GOT TO and 1 of GOT TO), the first of which is recorded in 1783. As the other options yield thousands of examples, I decided to exclude this marginal variant due to its lack of representation at the time (see Chapter 4).

Grammatical variables such as these investigated here were not the focus of early sociolinguistic work, which was mainly interested in phonology and the spoken

language. Interest in the morphological and syntactic level came later, often in historical studies, where the phonological level was not directly observable. In principle, the basic definition of any linguistic variable is "two or more ways of saying the same thing", as Tagliamonte (2012: 4) puts it in a recent textbook on variationist sociolinguistics. This seems straightforward. It gets interesting – and at times complicated and controversial – once we consider her follow-up question "What does it mean to say two things mean the same thing?" (Tagliamonte 2012: 4). Grammatical variables present a greater challenge than phonological variables when answering this question. Holmes (1994: 30) notes that alternative morphological variants like different past tense inflections such as e.g. BrE standard *-ed* and the zero morpheme, found in several varieties of English (Kortmann & Lunkenheimer 2013), are still relatively easy to set up, but that syntactic variables pose greater difficulty. It is, for instance, debatable whether a passive sentence and an active sentence can truly be considered to be two alternative ways of 'saying the same thing'.

The issue of equivalence of meaning concerning grammatical variables has been much debated in the variationist literature (see e.g. Lavandera 1978 or Sankoff 1988 for relatively early and influential discussions of the matter). The approach I subscribe to in this paper is based on Rydén's (1991) conception of grammatical variables, which he describes in a paper discussing the BE/HAVE variation:

> Variants in a syntactic (synsemantic) paradigm function in contexts above word or phrase level, i.e. as part of clausal/sentential constructions, with some kind of paradigmatic constant in structure and/or concept. (Rydén 1991: 345)

This 'paradigmatic constant', or common denominator of the variants in a paradigm, "may be difficult to pinpoint [...] in terms of neat definitions", Rydén (1991: 344) admits, as syntactic paradigms are "open-ended, complex and often semantically elusive". Some paradigms are primarily relational, such as the relative marker system, while others are primarily notional, such as modal systems (Rydén 1991: 345). It is part of the analyst's job to reflect on whether a case for variability between different options can be made, and under what circumstances this variability holds.

Another important aspect needs to be considered when investigating syntactic variables involved in change processes (as exemplified by the variable '(semi-) modals of obligation' in this study): syntactic change does not necessarily involve variant

addition or variant loss. Syntactic change can just as well be "'merely' a matter of systemic redistribution or change in markedness (in the form of 'specification' or 'generalisation') of the variants available, involving contextual refinements" (Rydén 1984: 515). In fact, Fischer (2008: 58) remarks that the introduction of a completely new structure or the complete replacement of one variant by another are only rarely found on the syntactic level. Much more frequently, we witness change as shifts in the relative frequency of particular constructions, the extension of structures to previously impossible contexts or their restriction to particular contexts (Fischer 2008: 58).

Finally, each of the four linguistic variables described above will also be impacted by a number of other linguistic variables. These will be reviewed based on earlier research in the chapters dealing with the individual features, and will be integrated into the analysis and the modelling as independent variables next to the social variables.

3.4.2 Social and other extralinguistic variables

The principal social variables under investigation in this study are gender and social class; a speaker's role in the courtroom is also included where useful. As a person's identity is not just the product of these two or three factors, it is clear that these variables can only paint a part of the picture. Any study has to deal with this issue. It is therefore extremely important that the social categories set up in a particular study can be shown to be meaningful for a particular context under investigation (see e.g. Bergs 2012: 88), and "on some level [...] *recognizable* to members of the community" (Kiesling 2013: 454). The present section will explain in how far the dimensions discussed in this study, listed in Table 7, are relevant to members of the Late Modern community and to Late Modern courtroom interaction.

Dimensions	**levels**
GENDER	male, female
SOCIAL CLASS	higher (HISCLASS 1-5), lower (HISCLASS 6-13)
ROLE	judge, lawyer, victim, witness, defendant

Table 7. Social factors (independent variables)

In addition to the dimensions listed in Table 7, the Old Bailey Corpus also provides information on scribes, printers and publishers of the issues of the *Proceedings*, wherever available.[44] These factors will be discussed when appropriate.

It is widely acknowledged in linguistic research that GENDER is socially constructed and performative.[45] Gender categories, most notably the distinction made between women and men, along with corresponding roles and behaviours are taught, learned and reinforced in everyday interaction – consciously and unconsciously. Considerable effort goes into constructing and maintaining gender differences, i.e. the social distinction between men and women. Linguists are well aware that "variation based on gender may not always be adequately accounted for in terms of a binary opposition" (Eckert 1989: 247). For instance, contrasting men and women as two large groups erases the multiple competing masculinities and femininities that people in these two groups represent and perform. While it is true that any society will have its hegemonic forms of masculinity or femininity that many adopt, there will always be people who do not conform to and do not identify with these. We need to assume that underlying the generalisation of 'woman' and 'man' are a variety of very different backgrounds, roles and expectations at any given point in time (Kiełkiewicz-Janowiak 2012: 313).

Having acknowledged that, it is nevertheless the case that sociolinguists usually conceptualise gender as a female-male dichotomy – and that this construct is useful in linguistic research. The present study, too, will use the categories 'men' and 'women'. This is not an endorsement of the gender binary as such, which is at best an oversimplification of people's realities and at worst oppressively normative to anyone not identifying along these lines.[46] Rather, the decision to conceptualise gender as a male-female dichotomy was made in light of the available information (annotation for 'male' and 'female' in the OBC), the methodological set-up of the study (quantitatively oriented) and the prevalent ideology of gender in Late Modern society.

Importantly, the gender binary was meaningful to Late Modern speakers, and it still is today: "The ubiquity of gender in publications about language variation is tied

[44] See Appendix B: Table B-1 for a list of scribes and printers, and Appendix B: Table B-2 for a list of publishers.
[45] The term 'gender performativity' was coined in Butler (1990) and has since been adopted in a number of academic fields, including linguistics.
[46] Research has also challenged the notion of supposedly 'biological' sex, with biologists demonstrating that sex is a spectrum. An accessible introduction to the most recent scientific discussion, published in renowned journals like *Nature*, is found in Ainsworth (2015).

to its status as a particularly salient, meaningful and cross-culturally relevant social category" (Queen 2013: 368). In particular, the aspect that is salient and meaningful is the binary distinction woman-man: children are, after all, socialised to be either one or the other. To this day, many people's main concern when it comes to gender is being able to distinguish between men and women and unequivocally assign gender to individuals. The insistence on the gender binary, imbued with differing expectations and norms of appropriate behaviours for men and women, is the reason why there is a robust pattern of variation by gender (understood as man vs. woman) since the beginning of the variationist endeavour (Queen 2013: 370).

Among the most well-known observations on gender-specific language use are the principles recorded by Labov (e.g. in Labov 1990) on the linguistic differentiation of women and men, which are considered some of the "clearest and most consistent results of more than 30 years of sociolinguistic research in the speech community" (Cheshire 2002: 425):

> Principle I: For stable sociolinguistic variables, men use a higher frequency of nonstandard forms than women. (Labov 1990: 210)
>
> Principle Ia: In change from above, women favor the incoming prestige form more than men. (Labov 1990: 213)
>
> Principle II: In change from below, women are most often the innovators. (Labov 1990: 215)

As Cheshire (2002: 426) points out, Principles Ia and II relate to language change, and Principle I refers to (temporarily) stable variation. It seems that "[w]omen conform more closely to sociolinguistic norms that are overtly prescribed, but conform less than men when they are not" (Labov 2001: 293).[47] At least, that is the case for contemporary speech communities. Good overviews of gender effects in variation and change can be found in Cheshire (2002) and Queen (2013), two handbook chapters set up to complement each other.

The investigation of gender in a historical dimension brings with it some additional issues – Bergs (2012: 89) admits that it "can be a methodological nightmare". A basic problem is that we cannot ask about speakers' gender identities

[47] Of course, whenever such a result manifests, it is necessary to check whether a 'gender pattern' is truly at the heart of observed variation. Cheshire (2002: 428), for instance, criticizes that this generalisation about women and standard forms "seems to be passing into the accepted sociolinguistic wisdom, without explicit recognition of the fact that statements involving class, prestige or 'standardness' are less objective than has been supposed".

but in fact are left with demographic data about sex (Kiełkiewicz-Janowiak 2012: 313). Also, we cannot assume that the linguistic behaviour of men and women in the past is the same that men and women exhibit in present-day society. Cameron (2008: 298) stresses the necessity to think about "whether the gender-linked sociolinguistic patterns most commonly reported in (Western) societies today are actually of rather recent origin, reflecting conditions which would not have obtained in most other times and places". In fact, these observed patterns are not a result of a person's gender per se but of what being (identified as) a person of that gender means in a given society in terms of roles, opportunities, expectations and (sometimes implicit) code of conduct.

For the Late Modern period, gender differentiation into men and women impacted everyday life and language use. Emsley et al. (2015g) state that "[v]irtually every aspect of English life between 1674 and 1913 was influenced by gender", arguing that "[l]ong-held views about the particular strengths, weaknesses, and appropriate responsibilities of each sex shaped everyday lives, patterns of crime, and responses to crime" and thus also directly impacted behaviour and reactions documented in the *Proceedings*. Sairio (2009: 46) also highlights the unequal status of women and men in 18th-century England: legal standing and educational opportunities were much greater for men. The idea of "separate spheres" for men and women is often invoked for this period, where the woman's domain is the home and the man is supposed to take part in public life. Although there was by no means total separation of men and women along these lines (Emsley et al. 2015g), it is certainly true that the growing "ideology of domesticity" (Price 1999: 41), especially prominent in the 19th century, pushed this separation and indeed closed many doors for women in public life. Some of the tendencies observed in this respect were by no means new: the continuing exclusion of women from skilled (and therefore better-paid) work in manufacturing and farming (e.g. by legislature and by a narrowing of what was considered 'women's work') and from public roles in business had already begun in the 17th century and was continued in the Late Modern period (Price 1999: 39–44).

Gender-specific expectations also carried over into recommendations for linguistic behaviour. In the 19th century, women were exhorted to listen carefully rather than speak, to use language in an appropriate manner when speaking and to avoid hurting the feelings of interlocutors (Kiełkiewicz-Janowiak 2012: 316). These

recommendations were based on the idea that women should be (or were 'naturally' thought to be) very sensitive to others' needs, responsive and agreeable (Kiełkiewicz-Janowiak 2012: 316) – an assumption that is still widespread today. The different roles and expectations placed on women and men were thus felt in an everyday way and led to a society in which relations within the group of women and within the group of men were more familiar than between people of different genders (Sairio 2009: 47).

SOCIAL CLASS, the second social factor covered in this study, is described with mixed feelings by Ash (2013: 350), namely as "universally used and extremely productive" in linguistic research but also as poorly understood by linguists. Ash (2013: 350) criticizes that class is often defined "fairly loosely" in linguistic studies, i.e. without building on the expertise of "other disciplines that make it their business to examine social class, particularly sociology".

Social class is composed of objective and subjective components, the first being the economic measures of property ownership and the second being measures of status (Ash 2013: 351). The "white collar" – "blue collar" distinction is the simplest way of applying this combination of objective and subjective measures (Ash 2013: 351). While social class is composed of several aspects, the most influential is occupation: in fact, occupation is "the single indicator that accounts for by far the greatest portion of the variance" (Ash 2013: 365). Many social class schemes are thus occupation-based. The HISCLASS system used in the OBC is no exception. It is based on the HISCO scheme, a classification system for occupations that was developed by social historians based on occupational titles found between the late 17th to the late 20th century (van Leeuwen et al. 2002: 13). Previous research has indicated that HISCLASS is suitable for Western "societies in the past three centuries, and possibly two more" (van Leeuwen & Maas 2011: 16), which makes it a good fit for the OBC material.

Class as a variable comes with some issues. In many early social dialect studies, women were classified according to their husbands' or their fathers' occupation. Romaine (1999: 174) draws attention to the problematic repercussions of this "patriarchal concept of social class", which considers the family, headed by a man, as the basic unit of analysis, and where the man's occupation determines the social

class of the family members (see also the discussion in Cheshire 2002: 428).[48] In a historical perspective, there is sometimes no other choice than to use such shortcuts, as we usually have less detailed information on women than on men and also because work outside the home was not always an option for women. The OBC compilers also made use of such strategies to assign class information to women (see Section 3.2.1).

The final sociolinguistic variable under investigation is ROLE: in the present study, lawyers, judges, witnesses, victims and defendants are distinguished.[49] A trial participant's role determines their behaviour to some extent, including their linguistic behaviour. Coding for roles like lawyer, judge, witness, defendant, etc. allows us to capture some of the power dynamics of the courtroom and also provides an opportunity to assess how courtroom dynamics interact with societal power dynamics on a larger scale (e.g. in terms of class or gender). Based on previous research (such as O'Barr & Atkins 1980, Archer 2005 and Cecconi 2012), it is reasonable to assume that different roles also come with differences in linguistic behaviour in the courtroom: it is conceivable that speakers with a lot of power (judges) and speakers in a very vulnerable position (defendants) may show differences, e.g. with regard to their degree of formality, or that 'courtroom professionals' like lawyers may have a different style than 'civilians' (witnesses, victims, defendants). It is necessary to underscore at this point that the variable role comes with some built-in issues that limit its analysis in the present work (see 3.5 for details).

It is important to remember that factors like gender, class and role are interrelated: one may determine the other (i.e. only men can have certain roles) or have an impact on how another is constructed (see the different experiences of female and male witnesses discussed in 3.2.4). It is not my aim in this study to discuss the different types of discourses and interactional styles of different groups, but it is worth keeping in mind that filling the same role as another person does not necessarily mean that you are having the same experience in court as another person with that same role

[48] It should be noted that the practice of assigning women and men to social classes in order to find out which men and women are of comparable status in a sociolinguistic study has been called into question in principle. Some researchers believe that women and men in the same social class are not actually of equal status, as the power balance is always tilted in favour of men in patriarchal societies, which "means that we can never compare like with like when we try to compare men and women" (Cheshire 2002: 428; also see Eckert 1989: 255).

[49] The roles 'interpreter' and 'jury' also exist in the OBC classification (see Table 7). However, they are not considered in the analysis as only very little material is produced by them. In addition, interpreters did not produce their own linguistic material, but interpreted for participants speaking a foreign language.

who happens to differ with regard to other characteristics such as gender or class. This is why possible interactions and intra-group variation need to be considered carefully.

Finally, the dimension of time, i.e. when an utterance was made, needs to be considered in a diachronic study such as this one. In the present work, time is operationalised as the categorical predictor PERIOD with either four (1720-1769, 1770-1819, 1820-1869, 1870-1913) or three levels (1710-1780, 1780-1850, 1850-1920).[50] This turned out to be the most practical solution for the present analyses and most useful in comparisons with results from other studies and between the OBC and the CLMET, as Chapters 4-7 illustrate.

3.5 Methodological concerns and analytical procedure

This section describes the general outline of all linguistic analyses in this study, which follows the structure shown in Table 8.

Analytical steps	
1	Extraction of tokens from OBC and CLMET; coding
2	Analysis of OBC results in terms of diachronic variation and change, focusing on the impact of sociobiographical factors
3	Comparison of OBC results with a CLMET subcorpus representing a different speech-related genre (drama or narrative fiction) in order to contextualise findings and assess 'external fit'
4	Research contemporary norms of correctness for the variable in question based on Late Modern grammars
5	Discuss variation and change in variables with regard to the interplay of factors mentioned in steps 2-4 (linguistic and extralinguistic factors) and against background of existing research

Table 8. Analytical procedure

As the features investigated raise very different issues in their analyses, though, each of the following chapters will have a brief section explaining the methodological approach for a given feature. In the present section, only a short overview of the process is provided. After that, central issues in a corpus-linguistic study such as coding, metrics and statistics will be touched upon.

[50] Note that in the three-period structure, the three periods are named as they are in the CLMET: period 1710-1780 actually only contains data from the years 1710-1779, and period 1780-1850 only contains material from the years 1780-1849.

The analysis has 5 steps. In step 1, the linguistic material is extracted from the corpora. For the OBC, the corpus-specific web interface was used, while the CLMET was searched with the free software AntConc (Anthony 2014). Which concrete search strategies were employed (e.g. whether POS-tags were used) depended on the phenomenon at hand. As a rule, the search patterns were designed in a way that privileged recall over precision. This increased the load of manual post-processing but was most useful for the variables in question. As part of the first step, the individual tokens were also coded for additional relevant linguistic variables: for the (semi-) modals, for instance, semantic coding (root – epistemic) was very important. Details can be found in the individual chapters.

In step 2, the OBC results are considered by themselves, but step 3 provides an important contextualisation of the results. Here, the data gained from one particular corpus, the OBC, is compared to another corpus. Either the drama subcorpus or the narrative fiction subcorpus of the CLMET were chosen for the comparisons. They are similar to the OBC in an important respect: they also contain what Culpeper & Kytö (2010) identify as speech-related texts. At the same time, they show important differences to the trials in the OBC. Drama is both speech-like and speech-purposed whereas trials are speech-like and speech-based.[51] Narrative fiction with dialogic passages is considered speech-like. Trials are based on actual utterances in real life while drama and fiction contain constructed dialogue. The corpora represent different genres with different genre conventions, which are united by the fact that they are an approximation of the spoken language (Culpeper & Kytö 2010: 3). Comparisons across corpora can provide a more accurate picture of overall linguistic trends and can also point out where idiosyncrasies of one genre are at play rather than global developments in language.

A further important consideration is made in step 4. The study's aim to describe the social profile of change and variation in the verb phrase encompasses also the consideration of the contemporary norms of correctness and whether they might interact in some way with the developments observed in the data. For the present study, Late Modern norms of correctness were established based on Sundby et al.

[51] Actually, a case can even be made that some trial proceedings are also speech-purposed in the sense that their authors deliberately made them look like speech so that they would be considered authentic and dramatic by readers (Culpeper & Kytö 2010: 17–18).

(1991), a dictionary of 18th-century normative grammars which gives a concise overview of grammarians' comments on individual features,[52] and a selection of 16 influential 19th-century grammars (listed by author, in alphabetical order, in Table 9).

Grammar (as listed in references)	**Edition used (if not first), and date of first edition**
Allen & Cornwell (1841)	---
Allen (1824)	3rd edition[53], "revised and improved" (first edition: 1813)
Beard (1854)	---
Bullen & Heycock (1853)	---
Crane (1843)	---
Crombie (1809)	2nd edition (first edition: 1802)
Curtis (1876)	---
Dawnay (1857)	---
Higginson (1864)	---
Hiley (1853)	5th edition, "considerably improved" (first edition: 1835)
Hort (1822)	---
James (1847)	---
Mason (1873)	18th edition, "revised and enlarged" (first edition: 1858)
Pinnock (1830)	---
Rushton (1869)	---
Turner (1840)	---

Table 9. 19th century grammars used in the present work

I am aware that any such selection can only represent a small section of the extensive grammar writing in the 19th century.[54] In the absence of a dictionary or database of 19th century grammars, though, I needed to establish a manageable list of grammars, which I could then canvass for comments on the variables in question. It was thus important to me to select grammars that would be very likely to comment on all four variables and cover the entire 19th century. I am grateful to Lieselotte Anderwald, who has worked extensively with 19th century grammars and created the Collection of

[52] For a critical assessment of the methodology and results of Sundby et al. (1991), see Anderwald (2016: 5–6).
[53] In those cases when I was not able to obtain the first edition of the grammar, I listed the one I used.
[54] Görlach's (1998) bibliography of 19th century grammars lists more than 2000 publications.

Nineteenth-Century Grammars (CNG),[55] for advising me on this matter and suggesting the grammars in Table 9. They cover a broad timeframe (1802-1876) and were all published in London. Two of the grammars on the list, Crombie (1809) and Mason (1873), also feature among the most influential grammars in the 19th century according to Wolf (2011), a monograph chronicling English grammar-writing between 1600 and 1900.[56] In addition to the comments retrieved directly from the grammars, information from previous work is included: grammatical opinion on the BE and HAVE perfect, for instance, is discussed in Anderwald (2012), Anderwald (2014b) and Anderwald (2016). Step 5, finally, synthesises all previously mentioned aspects of steps 1-4.

It should not be forgotten that a corpus-linguistic study makes assumptions that are to some degree idealisations. First, a corpus is assumed to be representative of a given genre or variety: the OBC is considered representative of spoken courtroom interaction in the Late Modern period, the CLMET-drama is taken as representative of plays and the CLMET-narrfic as representative of narrative fiction. Since these assumptions constitute a necessary methodological simplification of real-life complexity, it is even more important not to make sweeping statements about 'the English language' (whatever that may be) in the Late Modern period based on observations in just one genre. Comparison across several genres is needed to argue about more general developments. Secondly, genre is treated as a constant, i.e. it is assumed that all texts in the OBC between 1720 and 1913 represent examples of the same genre. In practice, we know that genre conventions and registers change (see 2.2.3, also Biber 2001).

As the present work is intended both to identify general tendencies and to serve as basis of comparison for individual's choices, some issues inherent in comparisons across studies need to be considered. The first issue in this respect is the construction of variables: as pointed out in 3.4, circumscribing a variable is, in the end, a judgement call. Two studies may include different variants of the same variable, for instance. The same problem applies to semantic coding, which is also employed in the present work. Of course, the rules applied to any semantic classification will be outlined in the

[55] In 2014, the CNG contained 258 grammars (Anderwald 2014b: 18). A list of all grammars contained in the collection can be found online: http://www.anglistik.uni-kiel.de/de/fachgebiete/linguistik/anderwald/cng-collection-of-nineteenth-century-grammar/cng

[56] Ten 19th-century grammars, four of which were published in England, are listed in Wolf (2011: 42). Influential works were defined as those that were re-printed and re-issued very often and thus in circulation for long stretches of time.

respective sections, but it remains a fact that the analyst decides how to structure the data by deciding which coding scheme to apply, how many categories to use, how to treat unclear examples (put into separate category, discard entirely, sort into existing categories to the best of their ability). Even if classification schemes are used that have been employed in previous studies by other scholars, difficulties with comparisons between results of different semantic analyses are ultimately "unavoidable" as "coding is a subjective exercise" (Aarts et al. 2013: 34).

Comparing results across studies is sometimes also made difficult by the use of competing measures of frequency and statistical measures. In connection with this, Mair (2009: 24) points out that the traditional metrics in sociolinguistics and corpus linguistics differ: while sociolinguists usually provide group-specific realisation rates, such as percentages of the realisation of variables as particular variants (e.g. that speakers opted for HAVE TO in 36% of all instances in which root obligation was expressed), corpus linguists often express their findings in absolute or normalised frequencies (e.g. stating that the frequency of MUST is 1.3 pmw and that of HAVE TO 8.3 pmw in a given corpus). As this study is primarily variationist and focussed on alternation between different variants, percentages will be used in addition to absolute figures for the variants in question. Normalised frequencies per million words will be used occasionally where the nature of the data prevents the use of percentages.

The use of 1 million words as a baseline for normalisation is not uncontroversial. Normalisation per a given number of words in running text (often per million words = pmw) is very widespread but not always the best choice. Instead, it is best to choose an "informative baseline", i.e. "a baseline that eliminates as much extraneous variation as possible" (Bowie et al. 2013: 65, also see Aarts et al. 2013), which a word frequency baseline cannot always live up to. Bowie et al. (2013: 65) illustrate this with the example of modal use: arguing that a speaker's choice to opt for a modal auxiliary represents a grammatical choice within a verb phrase and not within all words in a text, they show that counting modals per million words can mask a lot of extraneous variation as text categories can differ in 'VP density', i.e. different frequencies of VPs per million words. A rise in modals per million words can thus be due to a growing number of VPs in general or to an actual growth in modal use (Bowie et al. 2013: 67–68). To factor out that extraneous variation, they recommend using VPs

or even tensed VPs as baselines for studies of modals (Bowie et al. 2013: 64–68). Similarly, Smitterberg's (2005: 48) study of the progressive in English uses a so-called S-coefficient, which is a "percentage of all finite non-imperative verb phrases (excluding BE *going to* + infinitive constructions with future reference) that are in the progressive", in addition to a frequency baseline.

Knowing about the issues involved in using a 'traditional' frequency baseline, frequencies per million words are nevertheless reported at times in this study. After weighing the pros and cons of different approaches, I consider this practice acceptable for the present purpose. This is based on several reasons. First of all, this study focuses on variation, will thus mainly report absolute frequencies and percentages and discuss these frequencies only for contexts in which variation is possible. Frequencies pmw (or per another suitable baseline) will only be employed to provide supplementary information. Secondly, the corpora in the present study do not include texts of different text categories: the OBC exclusively contains trial transcripts, the CLMET-drama only plays, and the CLMET-narrfic only narrative fiction. The issue of differing verb density in different categories is therefore not a prime concern. As a third aspect, practical considerations carry some weight: using more sophisticated measures is very impractical without a parsed corpus. For instance, the DCPSE, the corpus used in Bowie et al. (2013), is fully parsed, which makes extracting VPs much more straightforward than for non-parsed corpora like the OBC. In addition, the DCPSE is much smaller than the OBC – it only contains ca. 800,000 words.

Frequency information as such is obviously of little value – it needs to be interpreted in an appropriate way. Using statistical analysis to support this process is an established procedure in linguistics. This study will make use of binomial logistic regression analysis, which is suitable for categorical dependent variables and several independent variables. The analysis will be performed with the help of R (Version 3.3.1, R Core Team 2016), a language and environment for statistical computing, and RStudio (Version 0.99.902, RStudio Team 2015), a user interface for R. The regression modelling and analysis will follow the procedure outlined in Levshina (2015: 253–290), consisting of fitting a model and selecting (independent) variables, identifying potential problems with the model (testing for possible interactions,

identifying outliers and overly influential observations, checking regression assumptions), modifying the model if necessary and finally interpreting its output.[57]

To start with, a backwards stepwise model selection is conducted to arrive at a suitable regression model. This means that first a model containing all likely factors is fit, i.e. one containing the extralinguistic factors gender and class and also any linguistic factors that may exert an influence on the distribution of the dependent variable. In the following, factors that do not add significantly to the explanatory power of the model are dropped in a stepwise fashion, i.e. one factor after the other.[58] Where interactions between independent factors were indicated by prior exploration of the data, models including interaction terms are also fitted. These, too, only remain in the model where they add significantly to its quality. The only argument for retaining a non-significant factor in the model would be that it is theoretically motivated. In some cases, it can be useful in a linguistic discussion to use "a more 'fleshed-out' model" (Tagliamonte 2006: 151), i.e. to also include factors in a statistical model that would normally be removed because they do not significantly improve its explanatory potential. Including them may however better explain the findings in context, e.g. because it can "show that certain internal or external categories pattern similarly" (Tagliamonte 2006: 151). Finally, as outliers and overly influential observations can negatively impact the model and lead it to over- or underestimate actual trends, the data will be visually explored for such observations.[59] They will be commented on in the analysis and removed when indicated.

In addition, it is necessary to check whether the assumptions of a logistic regression are respected:

> Assumption 1. The observations are independent (of one another).
> Assumption 2. The relationships between the logit and the quantitative predictors are linear.
> Assumption 3. No multicollinearity is observed between the predictors.
> (Levshina 2015: 271)

[57] The R packages *car* (Fox & Weisberg 2015), *effects* (Fox et al. 2014), *rms* (Harrell Jr. 2014) and *visreg* (Breheny & Burchett 2014) are used for the procedure. The regression models themselves are fit using either *lrm()* or *glm()*.

[58] This process is aided by the function *step()*, which selects a formula-based model (such as a logistic regression model) by using the Akaike information criterion (AIC), which is a measure of the relative quality of statistical models.

[59] Following Levshina (2015: 270–271), the function *influenceplot()* is used.

Assumption 2 poses no problem because the present study uses no quantitative predictors (see 3.4). It needs to be checked, though, whether assumptions 1 and 3 hold.

In corpus-linguistic studies, two major issues are at play concerning assumption 1, the independence of observations: potential priming effects and speakers' idiolects. Structural priming refers to speakers' tendency to reuse syntactic constructions they have recently produced or comprehended: experiments have, for instance, shown that speakers prefer to use a passive construction after being faced with prime sentences involving a passive construction (see e.g. Bock's (1986) influential study). The second issue, a speaker's idiolect, comes into play when one speaker contributes several data points, e.g. when a speaker contributes several instances of MUST and/or HAVE TO in a study of obligative modality.

Both of these things routinely happen in corpus studies of morphosyntactic variation, and yet – as pointed out above – most corpus-linguistic work tacitly assumes that speakers or writers make each of their linguistic choices independently from previous choices (Bowie et al. 2013: 92). This is obviously a simplification but there are factors that justify this approach in some cases, including the present one. First of all, studies with a linguistic (e.g. Bowie et al. 2013) as well as studies with a statistical focus (e.g. Levshina 2015) point out that any problems with the assumption of independence can be minimised by using large samples of many different texts and/or samples collected over a large span of time.[60] As far as the OBC data are concerned, it is a fact that some speakers provide more than one example and that we cannot exclude that priming may play a role. At the same time, though, the OBC texts were collected over more than 200 years. In addition, just one issue of the *Proceedings* contains many different trial accounts and therefore many different speech situations involving different participants. Going by the above statements, it is therefore justifiable to work with the assumption of independence.[61] Priming is not generally integrated into corpus analysis – mostly because it requires a lot of work to adequately identify, code and

[60] Levshina (2015: 254-257; 271), for example, uses binomial logistic regression to investigate the use of two different verbs in Dutch causative constructions based on a collection of newspaper texts spanning several years and considers the regression assumptions met. Bowie et al. (2013: 92) state that the assumption of independence is "less problematic for large samples made up of many texts".

[61] An alternative would have been to model the individual speakers as random effects. After careful deliberation, this was not done. In the OBC, speaker IDs are not 100% reliable, as shown in the discussion of generic speaker IDs (in 3.2.4) and as evidenced by the fact that IDs are assigned by individual trial, meaning that people appearing in more than one trial (or even in cases where a trial was split into two issues of the *Proceedings*) are listed with more than one ID.

model it. As the OBC is a corpus of almost 14 million words and the analyses in this study consequently involve several thousands of data points per variant, this degree of annotation was simply not feasible here.

For assumption 3 to be met, there must be no multicollinearity between predictors. Multicollinearity, also sometimes called simply collinearity, is present when there is correlation between two or more factor groups (Tagliamonte 2012: 124). Whether this is the case can be tested by calculating the model's Variance Inflation Factor (VIF) scores:[62] these should not exceed the threshold of 10 to exclude collinearity (Levshina 2015: 272). Additionally, one important step is taken at the start of the model selection process to avoid this issue: ROLE is not included as a factor in any model, as certain roles are correlated with gender and class. Specifically, judges and lawyers are exclusively higher-class men. The effect of role will be investigated in an alternative manner (i.e. not as part of the regression model) where appropriate.

The regression results are reported following the guidelines set in Levshina (2015), in a tabular format that shows the estimated regression coefficients (in log odds ratios[63]), their standard errors and *p*-values as well as 95% confidence intervals for the estimates. In addition to the regression model, I make use of effect plots (using the R package *effects*) to compute and illustrate the effects of individual predictors in the models. While the regression coefficients (and confidence intervals) in log odds ratios represent the likelihood of the second level of a binary dependent variable when all other predictors in the model are at their base level, the calculations underlying effect plots are different: they provide probabilities (not log odds ratios), and they do not assume that all other predictors (except the one currently in focus) are at their base level. Instead, they create displays for single effects based on fitted values of individual terms in a model (e.g. fitted values for higher class vs. lower class in a model that distinguishes two social classes). To compute these values, the values of other predictors (such as e.g. gender, age, frequency of a construction) are "fixed at typical values", i.e. "a covariate could be fixed at its mean or median, a factor at its proportional distribution in the data, or to equal proportions in its several levels" (Fox

[62] The *rms* package contains a function for this purpose: *vif()*.
[63] Log odds (or logit) are logarithmically transformed odds. They can range from – Infinity (the natural logarithm of 0) to Infinity, and are centred around 0. Negative log odds for an outcome in a binary choice show that this particular outcome is less probable than another one, positive log odds show that it is more probable (Levshina 2015: 261).

2003: 1). Discussing effects based on such effect plots can make a regression model more accessible and easier to interpret. That probabilities are provided is a major benefit in this regard as it cuts out the "mental arithmetic" that is required to interpret effects on the logit scale (Fox 2003: 3).

The concordance index *C*, rounded to two decimal places, is reported as a general goodness-of-fit statistic for the model. In addition to the regression analysis, I will also conduct and report on monofactorial analyses, i.e. how a factor in isolation affects the dependent variable, where it is helpful for understanding or illustrating a particular effect. Where the chi-squared tests are used in order to decide whether differences between samples of data / groups are statistically significant, a significance threshold of 0.05 is used. As a measure of association between an independent and a dependent variable, Cramer's *V* will be employed.

A few words of caution are warranted: first, linguistic data is generally not in a format suitable to most statistical procedures. This is typically worst in corpora like the OBC, where certain groups are underrepresented and may produce very little data. There are many NAs (empty cells), for which information was not available, and the data are often badly distributed. Therefore, the sets on which we can test are smaller than the sets of all retrieved examples, which constrains the analyses. In the following chapters, the number of observations N that is relevant for a particular table, diagram or test will always be mentioned for the information of the reader. In addition, we need to be careful not to overestimate the explanatory potential of statistical analyses. As straightforward as it is, it bears repeating: correlation is not causation. We can only observe correlation and hypothesise about causation, and eventually formulate interpretations of what the observed data and statistics mean for the speech community in context. As such, the role of statistical analysis in linguistic work is clear:

> In sum, there are two goals for finding the 'best' analysis for your data. On the one hand, you must be driven to find the best fit of the model to the data. […] On the other hand, you also want to explain (and demonstrate) how the variation is embedded in the subsystem of grammar as well as in the community.
> (Tagliamonte 2006: 151)

Just like frequencies, statistical analyses need careful interpretation to have any meaning, and are not an end in itself.

3.6 Summary and outlook: Researching Late Modern English

The present chapter has outlined the empirical and methodological basis of the study, beginning with a general discussion of the value of corpus-assisted sociolinguistic studies of language (3.1). Two subsections were devoted to a closer look to the main sources of data, the Old Bailey Corpus (3.2) and the Corpus of Late Modern English Texts (3.3). Section 3.4 introduced all the linguistic and social variables under investigation in this study, and Section 3.5, finally, outlined the analytical procedure employed to answer the research questions that were initially formulated (see 1.2) and are repeated below for convenience:

A. How do variation and change manifest themselves in selected morphosyntactic features in Late Modern English with regard to the timing of change (if present) and its social and linguistic factors? How do different speech-related genres compare?
B. How are the variants evaluated in grammars of the time (positive / negative / changing)? Is there a correlation between this evaluation and actual use?
C. How suitable are the *Proceedings of the Old Bailey* (and trial proceedings in general) for historical sociolinguistics? How close to the dialogue uttered in the courtroom can we assume these published transcripts to be? What needs to be taken into account (e.g. in terms of scribal/editorial interference) when basing linguistic analyses on trial proceedings?

The following Chapters 4-7 will contain analyses of the individual features, and Chapter 8 will synthesise these findings.

Based on what is known about the source materials at hand and the linguistic features under investigation, we can formulate several hypotheses at this point:

(a) A change is expected for the following variables: MUST > HAVE TO; BE perfect > HAVE perfect
(b) Stable variation is expected for *you was* / *you were* and *I says* / *I said*. In speech, both variables show variation in present-day English.
(c) The development of these features will differ in the two corpora, as they represent different kinds of speech-related texts.

(d) Heavily stigmatised variants are expected to be rare in trial transcripts due to the formality of the situation and the rather formal genre conventions (compared to other speech-related texts).

The validity of these hypotheses will be explored in the coming chapters.

4 Modals and semi-modals of strong obligation and necessity: MUST and HAVE TO

> WILLIAM PERRY. On 18th January, about 10 o'clock at night, I was at the Greyhound public-house, Webber-row, Waterloo-road—I live opposite Mr. Kerr, and he and the prisoner came there in search of me, and asked me to move some furniture—I said it was an insane time of night to do it, but I agreed—the prisoner went to the house with me, and put the furniture in the van—I objected to go with him on account of the time of night, and said that I would go in the morning, but **he said he had to go to Paris at 12 o'clock that night, and must go**—he helped me to load the goods, and I took them to Mr. Kerr's.
> (OBC, t18640229-357)

This chapter traces the development of two expressions of strong obligation and necessity in Late Modern English: semi-modal HAVE TO and 'core' modal MUST. During this period, several crucial changes take place in this subsection of the modal system: the use of the semi-modal HAVE TO increases immensely, and the construction MUST + HAVE + participle (*he must have given him the money*) gains ground (Biber 2004b). At the same time, MUST finally loses its past tense use as speakers increasingly resort to semi-modal constructions like *had to* to express past modality. All this makes the 18th and 19th centuries "a time period of intensive layering of forms" (Tagliamonte & Smith 2006: 348) and a particularly interesting era to investigate the social dimension of these developments that still shape current usage.

The present chapter starts with a brief introduction to the formal and semantic properties of modal verbs and related structures in general and to the particularities of the (semi-)modals discussed in this chapter (4.1). Existing research on English modality, especially with a focus on the obligation/necessity modals and on diachronic change, is surveyed in 4.2, before methodological challenges to the present study are addressed in 4.3. The chapter closes with a summary of the findings and their interpretation (4.4 and 4.5).

4.1 Introductory remarks

This section provides background information on modals and semi-modals in general and on MUST and HAVE TO in particular. Section 4.1.1 starts with a general introduction to the concept of modality and its grammatical encoding in modal and semi-modal verbs in English. Afterwards, a brief history of MUST and HAVE TO is provided in 4.1.2.

4.1.1 Modal verbs and related expressions of modality

The concept of modality comprises such diverse semantic notions as ability, possibility, hypotheticality, obligation, and imperative meaning, but all modal utterances have certain characteristics in common: "[They] are non-factual, in that they do not assert that the situations they describe are facts, and all involve the speaker's comment on the necessity or possibility of the truth of a proposition or the actualization of a situation" (Depraetere & Reed 2006: 269). Cross-linguistically, modality can be coded grammatically via a number of strategies, such as verbal inflections,[64] auxiliary verbs, adverbs or particles (Depraetere & Reed 2006: 270).

In English, modality is chiefly expressed by modal auxiliary verbs (Depraetere & Reed 2006: 270) and semantically related verbal structures which are usually called semi-modals.[65] Modal auxiliaries (also frequently termed modal verbs or simply modals) share a number of distinct properties. Some of those mark them as part of the larger group of auxiliaries (including e.g. also BE and HAVE) while others more narrowly delimit the group of modal auxiliaries.

Like all English auxiliaries, modals can occur in the four so-called 'NICE' constructions: "negation, inversion (of subject and auxiliary), code (post-verbal ellipsis dependent for its interpretation upon previous context), and emphasis (emphatic polarity involving the use of contrastive stress)" (Collins 2009: 12). Other aspects they share with all auxiliaries are their negative inflectional forms, (to a large extent) their phonologically reduced forms and their tendency to precede frequency adverbs, modal adverbs and quantifiers (Huddleston & Pullum 2002: 101–102).

Modal auxiliaries are set apart from other auxiliaries by additional properties: modals have no non-tensed forms, do not show person-number agreement and take

[64] Where verbal inflection is used to code modal meanings, we speak of mood (Depraetere & Reed 2006: 270).
[65] There is no clear consensus regarding terminology: instead of semi-modal, other terms such as quasi-modal can also be encountered in the literature.

only bare infinitival complements. Two further particularities are also of note, as for instance Huddleston & Pullum (2002: 108) point out: in unreal conditionals, the first verb of the apodosis needs to be a modal, and "[t]he preterites of the modal auxiliaries – *could, might, would, should* – can be used with the modal remoteness meaning without the grammatical restrictions that apply in the case of other verbs, where it is found only in a small set of subordinate constructions". Table 10 provides an overview of the above-mentioned properties of modals, each accompanied by one constructed example.

Auxiliary properties	
1) Primary verb negation	*She will not leave.*
2) Inversion of subject and auxiliary	*Must I do it?*
3) Code	*They won't do it, but I will Ø.*
4) Emphatic polarity	*He* COULD *help us, I'm sure.*
5) Negative inflectional forms	*You can't leave.*
6) Reduced forms	*He'll get over it.*
7) Precede adverb/quantifier	*His two remaining employees will probably both resign.*
Modal properties	
8) No non-tensed forms	**to can, *musting*
9) No agreement	*The patient will/*wills survive.*
10) only bare infinitival complements	*I must go /*to go.*
11) First verb in apodosis of remote conditional	*If you came over tomorrow, you could help me with the laundry.*
12) preterites can be used with modal remoteness meaning	*Could you change the booking?*

Table 10. Properties of modal auxiliaries (based on Coates 1983: 4–5, Collins 2009: 12–14 and Huddleston & Pullum: 92–115)

It has to be noted, though, that not every modal has all these properties:[66] for instance, MAY does not have a reduced form (criterion 6), and the negative *mayn't* (criterion 5) is unacceptable to most present-day speakers (Huddleston & Pullum 2002: 109). Moreover, different accounts of the English modals may not agree on the members and delineation of the category 'modal'. Despite these complications, however, the nine verbs CAN, COULD, MAY, MIGHT, SHALL, SHOULD, WILL, WOULD and

[66] A detailed discussion of the applicability of the 'modal properties' to individual modal verbs can be found in Huddleston & Pullum (2002: 108–115).

MUST are generally accepted as modals (Biber et al. 1999: 483–484, Coates 1983: 4–5, Huddleston & Pullum 2002: 92–115, Quirk et al. 1985: 136–148).

In addition to these central modals, grammars typically recognise a marginal or peripheral group of modals: Biber et al. (1999) as well as Quirk et al. (1985) consider DARE, NEED, OUGHT TO and USED TO members of this group. This is by no means undisputed: Huddleston & Pullum (2002) and Coates (1983) count NEED and DARE towards the central modal auxiliaries, and Huddleston & Pullum (2002: 92) further stress that USE belongs in a group of non-modal auxiliaries together with BE, DO and HAVE.

Beyond the central and marginal modals, classification becomes even more difficult. It is easy to demonstrate this with the help of Quirk et al.'s 1985 grammar: Recognising the plurality of forms and the difficulty to make rigorous distinctions between various types of verbal constructions with modal meanings, a gradient between modal auxiliaries at one end and full lexical verbs at the other is suggested (Quirk et al. 1985: 137). Six categories of verbs are distinguished, four of which are relevant to the discussion of modality: central modal, marginal modal (both discussed above), modal idiom (e.g. *had better*) and semi-auxiliary (e.g. HAVE TO) (Quirk et al. 1985: 137).

While the concept of a cline is indeed useful to account for the complexity among verbs with modal meanings, this four-part distinction is too fine-grained for the analysis at hand. For reasons of simplicity and efficacy, I will use a simpler two-part distinction between modals and semi-modals (as found in e.g. Biber et al. 1999). Modals, for the purposes of this study, are those nine verbs recognised as central modals by all four grammars mentioned above: CAN, COULD, MAY, MIGHT, SHALL, SHOULD, WILL, WOULD and MUST. The category of semi-modals includes the four marginal auxiliaries DARE (TO), NEED (TO), *ought to* and *used to* and the much larger group of "multi-word verbs which are related in meaning to the modal auxiliaries" (Biber et al. 1999: 73). Consider for instance HAVE TO, (HAVE) GOT TO, *(had) better* or BE SUPPOSED TO – all can code obligation meanings that share semantic space with the meanings expressed by modals like MUST or SHOULD.

This nicely illustrates the many-to-many correspondences between form and meaning that exist for modal verbs and semi-modals (Coates 1983: 26). [67] A single modal verb or semi-modal is usually equipped to express several meanings: MUST, for instance, is associated both with strong obligation (1) and with confident inference (2).

(15) You **must** do your homework.

(16) Alex **must** be here somewhere. I saw him enter this house.

According to Coates (1983: 24), the two meanings of MUST occur with relatively similar frequencies.

Modals and semi-modals are often sorted into groups by meaning. In Biber et al. (1999: 485), three groups are distinguished: (a) permission/possibility/ability modals (*can*, *could*, *may*, *might*), (b) obligation/necessity (*must*, *should*, *(had) better*, *have (got) to, need to*, *ought to*, *be supposed to*, (c) volition/prediction (*will, would, shall, be going to*). This list is not exhaustive (one could add colloquial forms like *gonna*, *gotta*), and other opinions exist on the dividing lines between groups or the number of postulated groups. For example, Coates (1983) distinguishes five groups of modals: obligation/necessity (*must, need, should, ought*); ability and possibility (*can, could*), epistemic possibility (*may, migh*t), prediction and volition (*will, shall*), and hypothetical modals (*would*, *should*).

More relevant to the present investigation than these differences between present-day classification systems is the fact that, whatever classification system is used, the groups cannot be expected to have been unchanging through time. Fitzmaurice (2002: 241) points out that the nine core modals, having completed the grammatical category shift from lexical to auxiliary verbs by around 1600 (see Rissanen 1999: 231–238), continued to exhibit semantic-pragmatic variation throughout the Late Modern English period. In addition to "sorting themselves into the deontic categories of prediction, necessity, and possibility, they gathered epistemic force, but these processes were by no means separate or ordered" (Fitzmaurice 2002: 241). Fitzmaurice summarises the development of the range of meanings available to individual modals in a tabular comparison (reproduced in Table 11) between the

[67] Due to their semantic flexibility, the modals and semi-modals have been conceptualised either as polysemous, i.e. possessed of several meanings, or as monosemous, i.e. having one basic meaning and other related meanings. An overview of influential polysemantic and monosemantic accounts as well as a model seeking to integrate the two can be found in Coates (1983: 9–22).

present-day state of affairs (as presented in Biber et al. 1999: 485) and usage around 1700.

	prediction/ volition	**obligation/ necessity**	**permission/possibility/ ability**
1999	will would shall	must should	can could may might
1700	will/would (3rd person)	will must shall should	can/could (not permission) may/might (permission)

Table 11. Categories of modal meanings in Early Modern and contemporary English (adapted from Fitzmaurice 2002: 241)

In the category of modals of obligation and necessity, which is the main interest in the present context, these changes are quite noticeable: Whereas *will* and *shall* belonged to this group around 1700, they no longer do. The picture is further complicated by changes to the system due to incoming semi-modals.

Another basic distinction in discussions of modality is that between epistemic and non-epistemic (or root) modality. Epistemic modality "reflects the speaker's judgment of the likelihood that the proposition underlying the utterance is true" (Depraetere & Reed 2006: 274). Generally speaking, this judgement can be anywhere between the two extremes of confidence and doubt (Coates 1983: 18). For instance, epistemic MUST in *he **must** be here* expresses a confident inference. Root modality, instead, "reflects the speaker's judgments about factors influencing the actualization of the situation referred to in the utterance" (Depraetere & Reed 2006: 274). Obligation and permission represent its core (Coates 1983: 21). *You **must** leave now* is an example of MUST expressing obligation, i.e. of non-epistemic/root MUST.

This root-epistemic distinction is by far not the only attempt to come to terms with modal meanings: many scholars have developed different categories and/or different cut-off points between them. Perhaps unsurprisingly, opinions can diverge quite radically between linguists based on their different theoretical preconceptions. Palmer (1990), for example, advocates a tripartite distinction between epistemic, deontic and dynamic modality, where epistemic modality is about speakers "making a

judgement about the truth of the proposition", and deontic modality is concerned with speakers "influencing actions, states or events" when giving permission or imposing obligations (Palmer 1990: 6).[68] Dynamic modality, finally, does not relate to the speakers at all, but "is concerned with the ability and volition of the subject of the sentence" (Palmer 1990: 7). Huddleston & Pullum (2002) also adopt a dynamic-deontic-epistemic distinction, but set different category boundaries than Palmer (1990). Quirk et al. (1985) distinguish between intrinsic and extrinsic modality (where intrinsic modality is characterised by some kind of human control over events, and extrinsic modality by the lack thereof). As an illustration of the plurality of approaches, those four examples shall suffice; more taxonomies exist, of course.[69]

Another distinction, cutting across categorisations such as root and epistemic, is that between subjective and objective modality.[70] Huddleston & Pullum (2002: 183) explain that for expressions of obligation the difference between subjective or objective modality is based on who or what imposes the obligation. Subjective modality is thus found whenever a speaker imposes authority on others or him-/herself (as in (17)). Where the authority comes from a source external to the speaker, though, we are dealing with objective modality (see (18)).

(17) You must clean up this mess at once.

(18) We must make an appointment if we want to see the Dean.

External sources are frequently rules and regulations.

Some scholars believe that certain modals or semi-modals express principally either subjective or objective modality: for instance, Huddleston & Pullum (2002: 206) claim that HAVE TO and HAVE GOT "characteristically differ from *must* in being objective rather than subjective", i.e. that MUST is used when the source of the modality is the speaker and that HAVE TO is largely reserved for external sources of modality. However, such generalisations are unconvincing in light of studies like Depraetere & Verhulst (2008), which have shown that "usage distinctions between *must* and *have to* are less clear-cut than reference grammars usually suggest"

[68] Following Lyons (1977: 823), Palmer (1990) conceptualizes deontic modality exclusively as relating to "the necessity or possibility of acts performed by morally responsible agents".
[69] For a brief summary of some of the most prevalent approaches and a helpful tabular comparison between these different schemes, see Depraetere & Reed (2006: 277–280).
[70] For an in-depth discussion of these concepts, see Lyons (1977: 797).

(Depraetere & Verhulst 2008: 23). A quantitative analysis of root necessity as expressed via HAVE TO and MUST in the ICE-GB reveals that both verbal expressions are used with a variety of source types such as "Speaker", "Regulation" or "Condition" – sometimes combined with each other (Depraetere & Verhulst 2008: 24).[71] Ultimately, it is not possible to establish a connection between sources of modality (and therefore subjective and objective modality) on the one hand and the type of (semi-)modal employed on the other hand.

Moreover, pinpointing the source of an obligation is often very difficult. Smith (2003: 242) remarks that his corpus study on root necessity left him with "very many instances where it [was] difficult to tell whether obligation [was] being reported or imposed, or agreed with". Finally, he decides against distinguishing sources of modality. Conceding that there may be characteristic associations of certain verbs with either subjectivity or objectivity, he ultimately agrees with Leech & Coates (1980) and Coates (1983) that "root necessity is a gradient phenomenon with no clear borderline between its intermediate stages" (Smith 2003: 242). In light of these issues, the present study will not distinguish between objective and subjective uses.

4.1.2 MUST and HAVE TO: a brief historical sketch

Many thorough and detailed histories of the English modals have already been written (examples include the discussion of the modals in Visser 1963-1973, Plank 1984 or Krug 2000). In this section, the aim is not to summarise them but to discuss key developments with regard to the study at hand. To this end, I will sketch the historical development of MUST and HAVE TO, including the evolution of their modal semantics, in broad strokes (a more detailed discussion of factors influencing variation between HAVE TO and MUST is found in 4.2) As the number of diachronic studies focussing only on HAVE TO and MUST is rather limited, I will also make reference to some works with a broader scope.

Of the modal expressions discussed in this chapter, MUST is the oldest. It has been available to express obligation "imposed from without, either by circumstances, regulations, legal prescriptions, etc., or by the will of a person since the end of the

[71] A preference for one of the verbs could only be detected for the source type "Circumstance", where users were partial to HAVE TO (Depraetere & Verhulst 2008: 24).

Middle English period" (Visser 1963-1973: 1805). Later, epistemic readings developed out of this root sense (Bybee et al. 1994: 195).[72] Molencki (2003: 81) dates this development to the late 14th century based on examples from the Helsinki corpus. The transition was probably eased by the presence of adverbs with a strongly epistemic meaning: Traugott (1989: 42) argues that "epistemic examples [of *must*] clearly expressing the speaker's assessment of the proposition first occur only in the environment of a strongly epistemic adverb, such as *nedes*". Molencki (2003: 81) especially points to the importance of the structure *must needs,* stating that "the epistemic sense was inferred from the adverb *nedes* (contemporary form: *needs*) rather than from the modal verb itself". Epistemic MUST only functioned without the support of *needs* or *necessarily* from the late 16th/early 17th century onwards (Molencki 2003: 82). Furmaniak (2011: 64–68) also argues that the co-text of MUST was instrumental in developing epistemic readings. While he makes no mention of adverbs like *needs*, he claims that two other constructions in particular promoted epistemic readings: MUST followed by state verbs (*he* ***must be*** *mad*) and MUST followed by perfect infinitives (*she* ***must have lost*** *her keys*). He proposes that these combinations created contexts that allowed for an indeterminate reading between 'inevitability', a type of root meaning, and 'probability', i.e. an epistemic reading. Using MUST in these contexts enabled the development of its epistemic sense, leading to a significant expansion of epistemic MUST in the 18th century (Furmaniak 2011: 55).

HAVE was available as a lexical verb denoting possession before it grammaticalised into an obligative semi-modal. The intricacies of this development have been discussed in a number of studies (Brinton 1991, Fischer 1994, Krug 2000, Lightfoot 1979, Warner 1993), partly using competing assumptions and explanations. What can be said with reasonable certainty is that the grammaticalisation process involved semantic and syntactic reanalysis of the construction 'HAVE + object + *to* + V' to 'HAVE *to* V + object' (see e.g. van der Gaaf 1931: 180–188, Visser 1963-1973: 1474–1487, Krug 2000: 53–61). This process is generally conceptualised in four stages (Brinton 1991: 10–11): In stage 1, HAVE is a full verb with possessive meaning. Stage 2 sees the coexistence of meanings of possession and obligation or duty, before the

[72] It is assumed that epistemic meanings of modal auxiliaries developed out of root/deontic meanings: for an argument based on evidence from child language acquisition, the history of English and a creole language, see Shepherd (1982).

possessive reading is bleached in stage 3 and HAVE becomes a modal auxiliary. In stage 4, use of auxiliary HAVE is extended to intransitive infinitives. While this sequence is generally accepted, the timing of these stages is a matter of debate (see Brinton 1991 for a helpful summary of the discussion). Most accounts date the appearance of HAVE TO with an obligation reading to Middle English (van der Gaaf 1931: 182–184, Visser 1963-1973: 1478), although it was probably still rather rare then.

What complicates matters is the fact that the original full verb HAVE + object + infinitive construction with possessive meaning split into two separate constructions with modal meaning in Early Modern English, i.e. *have to* V (+ object) and *have* + object + *to* V while the original construction also remained available (Brinton 1991: 27). Brinton (1991: 25) argues that there is a semantic difference between them: while the contiguous construction (19) emphasizes the duty to perform an action, in this case of writing a letter, the discontinuous construction (20) emphasizes the obligation to accomplish a result, in this case to have a written letter:

(19) I have to write a letter.

(20) I have a letter to write.

The structures behave differently under negation, and their syntactic bracketing also differs: In example (20), "*have* is less fully auxiliated, […]; its meaning is less restricted, encompassing meanings of possession as well as obligation" (Brinton 1991: 25). In the present study, I will only be concerned with the contiguous construction (19) that carries only obligation reading.

Krug (2000: 74) points out that obligative contiguous HAVE TO remained rare throughout the Early Modern period, and ambiguous readings were frequently found as typical complements with the verbs *say* and *do*. A significant increase in HAVE TO as an alternative to MUST can therefore only be observed in the Late Modern period:[73] based on ARCHER data, both Biber (2004b: 207) and Krug (2000: 74) report a gradual increase of HAVE TO around 1800 and then real growth in the 19th century, which prompts Krug (2000: 76) to declare that "quantitative research [of HAVE TO] makes sense only from Early Modern English onwards" (Krug 2000: 76). Biber

[73] For a different view, see Tagliamonte & Smith (2006: 348), who assume that HAVE TO and MUST competed for obligative function from the Middle English period onward.

(2004b: 207) reports that the first such instance in the ARCHER letter corpus dates from the late 18th century.

The Late Modern period is thus characterised by a dramatic rise in the use of HAVE TO and in the expansion of MUST *have* + participle. From a present-day perspective, MUST cannot serve as an alternative to MUST *have* + participle. The lack of a past form of MUST in contemporary English is almost a staple element in discussions of the English modals (see e.g. Coates 1983: 40, Palmer 1990: 79, Quirk et al. 1985: 128). It is what makes its "paradigm incomplete even by modal standards" (Denison 1998: 176). In Late Modern English, the situation was not so clear-cut, as example (21), with epistemic MUST in a past context, illustrates. Root MUST in a past context is shown in (22).

(21) About Friday was Fort-night, I lost my Watch in an Alley in Chick-Lane, between 7 and 8 o'Clock at Night. The Prisoner **must** take it, because she and I had been in Company together […]. (OBC, t17360115-29)

(22) Q. How came you to go home with this woman that had robbed you of your watch [...]?
A. I was to go to pay for the coach.
Q. You did not hire the coach?
A. No; but I **must** pay three shillings, and she **must** pay three shillings.
(OBC, t17970920-24)

While it is true that MUST as a past tense was "virtually lost" in the Late Modern period,[74] leading to a "significant recent change" in the paradigm of this modal (Denison 1998: 177), this was a gradual process, and we can expect to find remnants of the old use.

The beginning of the process is best explained by going back to Old and Middle English: "In origin *must* was a past tense (OE PRES 3 SG *mot*, PAST 3 SG *moste* 'be allowed to, may'), but over the course of the ME period it came to serve also as a present tense" (Denison 1998: 176). This development was accompanied by semantic changes: in Old English, *mot* generally expressed permission and possibility but also acquired dynamic and deontic necessity senses (Warner 1993: 160). In ME, its use as a marker of permission was greatly restricted, much of its functions being taken

[74] Some scholars argue that the disappearance of past indicative MUST happened earlier: For Krug (2000: 96), MUST has been a "past tense form without past time reference" since Middle English.

over by *may* (Warner 1993: 176). Instead it developed into a modal of necessity and obligation (Warner 1993: 176). When MUST, originally a past tense of *mot,* came to be used with present-time reference in ME, it was established as a separate verb from *mot*, which did not survive past the 16th century (OED[75] *mote* v.[1]).

The past tense use of MUST existed side by side with its present tense use for a time, but was increasingly reduced (see e.g. Rissanen 1999: 235 for the increasing replacement of such modals by periphrastic expressions like *had to*, *wished to*, etc. in Early Modern English). In contemporary English, the use of MUST to express necessity or obligation in the past is "mostly confined to instances of oblique narration, and of the virtual oblique narration in which the speaker has in his mind what might have been said or thought at the time" (OED *must* v.[1], II.2). Denison (1998: 176) also mentions its survival in "certain backshifting contexts". These backshifting contexts, where a past form is required in the reported clause because the verb in the superordinate reporting clause is in the past tense, are also found in Late Modern English, as illustrated in (23).

(23) JAMES WILLIAM CROUCH. (Policeman, R 118). [...] Mr. Knight called the prisoner into the parlour. I told him he **must** take his apron off, and go with me, as I was a policeman [...]. (OBC, t18570615-742)

There is no consensus on the status of MUST in such contexts, i.e. whether it is truly a surviving past-tense form of MUST or a present-tense form substituting for a past tense that is no longer available. Denison (1998: 178), for instance, argues that examples of MUST in past contexts from the late 19th / early 20th century onwards are not "relics" of a past tense MUST but "early instances of a present tense modal in a past tense context" that form "part of an incipient loss of the general backshifting rule […], at least as far as modals are concerned".[76] For the present purpose, I make no distinction between MUST in backshifting contexts and MUST in non-backshifting contexts, as both uses are

[75] 'OED' here refers to the online version of September 2017, which can be found in the list of references under "Oxford University Press (2017)". When other (print) versions of the OED are cited, this is explicitly mentioned.

[76] Other scholars also argue against MUST being a past tense form in oblique narration and similar contexts: Palmer (1990: 121) considers the use of MUST in such contexts not as evidence for the survival of past tense MUST, but simply as a result of the sequence of tense rules: "MUST is used as if it were a past tense form, to report present tense MUST in any of its uses". It is simply a present-tense form substituting for an unavailable past-tense form. More generally, Palmer (1990: 121) argues that "past tense forms are needed by the sequence of tense rules, and not to indicate the past time of the event" and is adamant that MUST "has no past tense form". Jacobsson (1979: 303) shares the view that "MUST cannot by itself indicate past time", but that "'speech' […] in the widest possible sense including thoughts, rules, regulations, and the like" can legitimise "the use of MUST in past tense environments". Coates (1983: 40) and Quirk et al. (1985: 128) argue in a similar vein.

essentially uses of MUST in a past context – and thus potential alternatives either to *had to* in root contexts and potential alternatives to MUST *have* + participle in epistemic contexts for Late Modern English.

4.2 Previous research: variation and change among the modals and semi-modals of obligation and necessity

This section reviews literature on the diachronic development of English modals and semi-modals in general (4.2.1) and on expressions of obligation and necessity, especially MUST and HAVE TO, in particular (4.2.2). The Late Modern period is the focus of attention, but studies covering a longer period of time as well as some explorations of 20th century trends (4.2.3) will also be taken into account. The aim is to bring together those factors that exert an influence on the choice between HAVE TO and MUST. To simplify the comparison and contextualisation of the different studies' results, the discussion is primarily structured by period and by the corpora used in the analysis. Finally, Section 4.2.4 provides information on how contemporary grammars viewed the use of the alternatives under consideration.

4.2.1 Modals and semi-modals in a diachronic perspective

Researchers generally agree on two major developments in the area of English modality: 1. the modals have been in decline for about a century. 2. The semi-modals are increasing in use. However, there is considerable debate on whether these processes are linked. This section summarizes major developments in the English modal system and considers the relationship between modals and semi-modals based on corpus studies undertaken in this area.

A long-term perspective on English modals as a group is provided in Biber (2004b). Using parts of ARCHER and additional material from the Longman-Lancaster and BNC corpora, the development of English modals and semi-modals is investigated from the 17th century onwards. The results show that the modals as a group have sharply declined in the last 50-100 years, whereas the use of semi-modals has increased (Biber 2004b: 199). This trend is corroborated in a number of studies with a focus on the 20th century, most of which are based on the Brown family of

corpora. Leech (2003) reports a decline in the frequencies of modals as a group between the 1960s and the 1990s based on the Brown quartet. A follow-up study (Leech 2011), which draws on additional British corpus data from around 1900, 1930 and 2006, and COCA and COHA for American English), corroborates these results. Leech (2011: 561) concludes that "the frequency decline (in standard AmE and BrE) of the modal auxiliaries as a class is now past reasonable doubt". At the same time, an increase in semi-modal use has been observed (e.g. Leech & Smith 2006: 188). In light of such findings, one of the questions that emerge is whether the rise of the semi-modals is in some way connected to the decline of the modals, and if yes, in what way.

An important consideration in this respect is the fact that the semi-modals are available in contexts where modals, due to their incomplete verbal paradigms, are not. The semi-modals, after all, serve to "fill gaps created by the peculiar morphology and syntax of the modals" which have no tensed or non-finite forms (Palmer 2003: 15). At this point, it is important to differentiate between two different scenarios: either that the defective paradigm of the modals caused the rise of the semi-modals, or that the rise of the semi-modals was strengthened by the availability of such syntactic gaps left by core modal morphology.

Lightfoot (1979: 112) proposed a causative explanation, according to which the semi-modals entered the language for the express purpose of compensating for the incomplete paradigms of the modals. The modals had only shortly before emerged as a new grammatical category in late Middle English due to a number of drastic changes in the verb phrase. This account, often dubbed a 'catastrophe scenario', has been widely rejected in favour of more gradual explanations (e.g. Plank 1984, van Kemenade 1992 or Warner 1993). Most researchers adopt the view that the syntactic flexibility of semi-modals may well have aided their spread, but do not assume a causal link like Lightfoot (1979). Görlach (2001: 123–124), for instance, claims that semi-modals became more frequent in the 18th century partly because they "could be used in non-finite forms in syntactic frames where the core modals were impossible". Krug (2000: 95) also considers it likely that "the spread of the quasi-modals is connected with the failure of the central modals to occur in certain contexts".

One argument against a causal link between the emergence of the semi-modals and the modals' defective paradigms is based on the contexts in which semi-modals

first occurred: Krug (2000) is able to show, based on ARCHER-I fiction and drama data, that semi-modal HAVE TO first occurred in contexts where its modal alternative MUST has always been available and still is:

> The fact that present tense forms are among the first attested uses of modal HAVE TO musters further evidence against a causal link between the genesis of the construction and the defective paradigm of the central modal verbs, because it is precisely such present environments where the modals have always been available. (Krug 2000: 95–96).

Dollinger (2006: 300) confirms this observation for HAVE TO and MUST in Canadian English: the first attestations of semi-modal HAVE TO between 1776 and 1799 were all in the present tense. Myhill (1995: 166–167), however, reports that in American English, HAVE TO first appears in what he calls syntactic contexts, i.e. environments where HAVE TO "is basically obligatory because *must* and *got to* cannot be used: lack of obligation (*you don't have to*), following a modal (*I'll have to*), past (*I had to*), and participal (*having to*)".[77] The picture is thus mixed.

Researchers are also sceptical about a causal link in the opposite direction, i.e. about the emergence of semi-modals causing the decline of the core modals. Leech (2003: 235) sees "little evidence that the use of semi-modals is a direct causative factor in the gradual demise of the 'true' modals", mainly because there is "no clear overall picture regarding semi-modals": many have been increasing since the mid-20th century, some have been declining, and most of them are much less frequent than modals. In fact, many corpus studies showed that the increase in semi-modals falls far short of making up for the decline in core modals. Citing ARCHER-I data on newspaper and academic prose, Biber (2004b: 199) states that the decrease in modal use "is not offset by a corresponding increase in semi-modal use", which he considers an indicator "that the two trends (decreasing modal use vs. increasing semi-modal use) are at least partially independent" (Biber 2004b: 199). For 20th-century English, Leech (2013: 96) remarks that "the overall frequency of core modals is several times that of the emergent modals" and that "the core modals are declining proportionately faster than the emergent modals are increasing".

[77] Contexts where HAVE TO is not syntactically motivated are called 'nonsyntactic contexts' (Myhill 1995: 166).

Important factors in such discussions of modals and semi-modals are language norms and issues of prestige. Leech (2013) argues that colloquialisation plays an important role in the general rise of semi-modals and the decline of modals, and also accounts for the presence (or absence) of semi-modals in writing:

> […] grammaticalization of the emergent modals in speech has been associated with increasing frequency, progressively leading to competition with the core modals, which consequently have been undergoing decline in recent English. Through colloquialization, the rise in emergent modals has been gradually filtering into the written language, but this process involves a time lag, and is probably impeded by a "prestige barrier".
> (Leech 2013: 114)

The "rather limited increase in emergent modals" (Leech 2013: 110) in writing is due to a 'prestige barrier' that prevents semi-modals from being used in writing as much as in speech. Of course, this has consequences for historical analyses of semi-modals in written sources. There is an undeniable gap between spoken and written registers, although a narrowing of that gap has been taking place for some registers like drama, fiction and press language (Krug 2000: 88). In these, we are more likely to find earlier examples of semi-modals, and a smaller gap between semi-modal increase and modal decline.

In addition, observations on general trends are by no means universally applicable across regional varieties, text types / genres or individual modals and semi-modals. Biber (2004b: 199) finds that the pattern of increasing semi-modal and decreasing modal use is stronger in American English than in British English and mainly restricted to drama and letters in both varieties. Millar (2009: 199) reports an increase of more than 20% in modal use between 1923 and 2006 in the 100-million-word TIME Magazine corpus – against the prevailing trend of modal decline. It is likely that this is a genre-specific trend, as Leech (2011) argues in response to Millar (2009). What Millar (2009) and Leech (2003, 2011) agree on is that not all modals follow the same downward trajectory: both argue that very infrequent modals are most affected by losses in frequency (e.g. *shall*), whereas very common ones persevere or even increase in use (e.g. *can*).[78]

[78] Millar and Leech do not completely agree on which modals should be considered members of which of group: MAY, for instance, is considered a modal in decline in Leech (2003), but on the rise in Millar (2009).

To summarize, the relationship between the modals and semi-modals is far from simple. For Biber (2004b: 211), the fact that different genres show different developments is proof that more than a "simple grammatical reorganisation (with modal verbs being replaced by semi-modals)" is taking place. Myhill (1995: 199), who does not exclude the possibility that "there may be a long-term structurally motivated process that will eventually result in the elimination of all of the true modals", points out that "this is clearly not a unitary process; it interacts with other factors, resulting in the elimination of some modals before others" (Myhill 1995: 199). In the end, not all modals and semi-modals behave alike. Many linguistic and extralinguistic factors, such as modal meaning (root – epistemic), the type of construction or speakers' social backgrounds may also have an impact on any of them. To achieve an overview of relevant factors for the investigation of MUST and HAVE TO, the next section reviews prior research with a particular focus on these modal expressions.

4.2.2 Long-term diachronic change in the domain of obligation/necessity

Diachronic corpus research with a focus on the modals and semi-modals of obligation and logical inference is still relatively scarce for the Late Modern period. Most historical studies are either more concerned about the big picture and the modals as a group (see 4.2.1) or, alternatively, focus on very particular settings and questions, usually based on smaller samples of text.[79] However, there are insightful studies tackling the obligation/necessity group in the Late Modern period (or at least allocating a decent amount of space to it within a broader investigation), which will be presented in the following.

For a long-term perspective on the domain of obligation and necessity, we can turn to a case study in Biber et al. (1998: 205–210), which discusses the development of *must, should, have to, got to, ought to, need to* and *supposed to* from the 17th century onwards based on several corpora (ARCHER, Longman-Lancaster and BNC). The results indicate that "modals have generally been more common than semi-modals over the last four centuries", but that this relationship is shifting, with "semi-modals becoming increasingly common" (Biber et al. 1998: 206–208). At the same time, these

[79]Examples of the latter type of study are e.g. Fitzmaurice (2002), analysing selected modals' pragmatic meanings in 18th century patron-client correspondence, or Nurmi (2013), investigating deontic modality in 16th century merchant letters as a means of negotiating power and social distance.

general trends cannot account for the behaviour of all modals and semi-modals. Despite the general increase of semi-modals, NEED TO and *ought to* are "relatively rare across all periods" (Biber et al. 1998: 209). HAVE TO, in contrast, has seen the most dramatic increase among all the semi-modals in the sample (Biber et al. 1998: 208–209). Register also plays a role for the frequencies of different (semi-) modals: For instance, "*(have) got to* shows a striking difference in its register distribution from *have to"*: while HAVE TO is prominent in all registers (news, fiction, drama/conversation), (HAVE) GOT TO only attains great frequency in the drama/conversation data and thus appears to be "restricted primarily to spoken English" (Biber et al. 1998: 209). Apparently, there is a considerable 'prestige barrier' for (HAVE) GOT TO (Leech 2013: 110, see also 4.2.1).

Biber (2004b), another long-term perspective on the English modal system, includes a section called "must vs. have to". Biber (2004b: 206–210) investigates the use of MUST and HAVE TO in personal letters from the 18th to the 20th century and notes that the frequency of MUST has remained relatively stable, but its range of "typical meanings" has expanded from predominantly expressing root to expressing both root and epistemic meanings:

> In seventeenth- and eighteenth-century letters, *must* almost always expressed meanings of personal obligation […]. In contrast, *must* in the twentieth century has extended its typical meanings to include both logical necessity and personal obligation meanings[.] (Biber 2004b: 206–207)

At the same time, "*have to* has been encroaching on the semantic domain of *must* in personal letters, being used more frequently to express meanings of personal obligation" (Biber 2004b: 208), [80] thereby taking ground from MUST in its core meaning of obligation. Biber (2004b: 207) does not provide absolute frequencies in his article, but cites four examples of "logical necessity" MUST to illustrate its rise in that domain. He does not address this directly, but I find it striking that three of those contain the construction *must have* + VVN and that the fourth example is *must* + *be*, i.e. those contexts where epistemic readings supposedly first appeared (Furmaniak 2011; also see Section 4.1.2). This is why I think the process is most accurately

[80] Not all of the examples mentioned in Biber (2004) would be considered examples of root modality (obligation) in the categorisation used in the present analysis: *have to acknowledge* (Biber 2004b: 208) would be classed as a performative use, for instance (see 4.3).

described as a modal extending not only its typical meanings but also its structural options, one of which becomes specialised for the new meaning. In fact, the construction *must have* + past participle even becomes obligatory for the marking of preterite tense for epistemic MUST from around 1700 onward (Molencki 2003: 85). An example is presented in (24).

(24) The prisoner **must have** gone inside (OBC, t18490226-669)

It can be assumed that this expansion in the formal and functional sense is responsible for the observed stable frequencies of MUST (Biber 2004b: 208), as MUST cedes ground to HAVE TO in its root sense. Once again, different genres exhibit discrete developments that may not fit the general pattern. In newspaper prose, for instance, MUST is increasingly restricted to expressing obligation, and it actually increases in use (Biber 2004b: 209).

A book-length investigation into the English modal system, which devotes a chapter exclusively to the emerging modals of obligation and necessity HAVE GOT TO (incl. *gotta*) and HAVE TO (incl. *hafta*) is Krug (2000). Data from the drama and fiction subsections of ARCHER (1650-1990s) show that both periphrastic modals increase throughout the period under investigation (Krug 2000: 80–81). As their "critical grammaticalization stage[s]", Krug (2000: 80–81) identifies the middle of the 19th century for HAVE TO and the early 20th century for HAVE GOT TO based on a drastic rise in textual incidence. Krug (2000: 80) suggests that HAVE TO "originated in discourse" based on the observation that speech-based registers like drama were quicker to accept the incoming forms than genres like fiction. He argues that the idiomatic expression *have to say/tell* was essential in the modalisation of HAVE TO (Krug 2000: 97):

> One might argue […] that HAVE TO *say* denotes an abstract type of possession, i.e. 'being in possession of, for instance, news or ideas '[…]. This, then, could be seen as a logical intermediate step in the progression from a possessive to a deontic reading.
> (Krug 2000: 101)

According to this explanation, *have to say* generates an agent-oriented reading: as the usual possession meaning is not suitable in the context, hearers interpret it as a deontic reading due to pragmatic inferences (Krug 2000: 101). In another step, this reading is generalised from verbs of saying to all verbs (Krug 2000: 102). Throughout the period under investigation (roughly 1650-2000), HAVE TO remains the more frequent semi-

modal (Krug 2000: 79). Differences in terms of regional varieties can be observed: American English was and is a forerunner in the spread of both semi-modals (Krug 2000: 79).[81]

A study that exclusively considers modality in American English is Myhill (1995). The author traces the development of numerous modals and semi-modals in the 19th and 20th centuries based on nine plays written between 1824 and 1947 and a collection of comics from 1984. In the obligation/necessity domain, *must, have to, got to, should, ought*, and *better* are considered. He claims that a dramatic change is noticeable in frequencies and functions of modals before and after the American Civil War (1861-1865): *must* and *should* declined sharply after the Civil War, and "other forms with the same general functions correspondingly rose in frequency — *have to* and *got to* for strong obligation, *better* and *ought* with weak obligation" (Myhill 1995: 159).

For Myhill, this is no simple replacement process. Instead, the "new" semi-modals differ from the "old" modals in a given semantic field (such as obligation/necessity) in terms of their subfunctions. He argues that the old modals expressed "principled" functions, i.e. functions involving "a clear social order and absolute evaluations based upon ostensibly universal principles", whereas the modals gaining ground after the Civil War expressed "interactive functions", i.e. functions that "presuppose more or less equal power relationships between people and focus on interactive factors such as mutual cooperation, emotional appeals, advice, apologies, or threats" (Myhill 1995: 160). For the semantic field of strong obligation, this development is summarised as follows:

> […] *must* was most typically related to social norms (e.g. *If he has committed a crime, he must be punished),* while *got to* is typically associated with emotional necessity (e.g. *You've got to help me!)* and *have to* is typically associated with habitual obligations (e.g. *He has to take the bus to work every day).* (Myhill 1995: 163)

Myhill (1995: 163) acknowledges that MUST is occasionally used to express emotional necessity or habitual obligations, and GOT TO and HAVE TO in association with social norms, but that "this is comparatively rare".

[81] For HAVE GOT TO, Krug (2000: 78) even suggests that it was an American innovation.

The decline of MUST and the rise of HAVE TO and GOT TO are thus results of speakers' changing needs to express different subtypes of modal meaning (Myhill 1995: 200). While the functions of modality that can be expressed in English did not change, speakers' need for using different subfunctions did (Myhill 1995: 159–160). MUST was frequent in the Antebellum period because its focal function, i.e. an obligation related to social norms and clear power structure, was especially relevant in that society (Myhill 1995: 159–160). The functions of habitual obligation and emotional necessity, respectively associated with HAVE TO and GOT TO, gained importance in the post-war period due to societal change (Myhill 1995: 159–160). The decline of MUST and the rise of alternative periphrastic modals are thus interpreted as symptoms of cultural change. A different idea relating to cultural change is expressed in Biber (2004b: 211), where the shifting frequencies of modals and semi-modals are seen as part of ongoing developments in the larger domain of stance expressions. These stance expressions are on the rise: "speakers and writers are apparently more willing to express stance in recent periods than in earlier historical periods", which is indicative of "a general shift in cultural norms" (Biber 2004b: 211).[82]

That both variety-specific developments and general trends are at work is shown by Dollinger (2006), who studies HAVE TO and MUST in American, British and Canadian English in the late 18th and early 19th centuries. At that time, "all varieties of English were headed towards more uses of *have to* at the expense of *must*", but variety-specific trajectories can nevertheless be identified (Dollinger 2006: 299). In particular, American usage was most progressive, followed by Canadian usage and finally British usage (Dollinger 2006: 295–301). It is remarkable that Canadian English aligned with American English despite an ongoing armed conflict between the countries and widespread anti-Americanism in Canada in the early 19th century.[83] For Dollinger (2006: 300), this constitutes evidence that the change in preference from MUST to HAVE TO occurred without social awareness, i.e. was a change from below in Labovian terms (for a detailed definition, see Labov 2006: 206–207). This is in line with what Krug (2000: 254) stipulates for British and American English: the change in preference towards HAVE TO is a change from below the level of awareness. What also

[82] Biber (2004a) explores historical patterns of stance marking in more detail.
[83] Dollinger (2006) refers to the War of 1812, a military conflict between the United States of America and the United Kingdom, its North American colonies (part of which now form Canada) and its Native American allies. The conflict lasted from 1812 to 1815.

points toward a change from below is that the first Canadian English occurrences of obligative HAVE TO are found in letters, i.e. rather informal texts (Dollinger 2006: 296).[84]

4.2.3 Recent developments: the 20th century

Although the 20th century is not our concern in this study of Late Modern morphosyntax, it is necessary to include more recent trends to contextualise our analysis. Thanks to the availability of several well-known corpora, especially the Brown family, there are many detailed studies on English modals of obligation and necessity for the (second half of the) 20th century. They largely support the long-term trends regarding (semi-) modal frequency identified in 4.2.1 and 4.2.2: the decline of MUST and the rise of semi-modal alternatives. However, they also help identify further variables that affect the use of MUST and HAVE TO. Another major advantage of more recent studies is the possibility to include real spoken data.

The decline of MUST in written contexts is confirmed for the recent past in both British and American English based on the Brown quartet of corpora (e.g. Leech 2003, Leech & Smith 2006, Leech et al. 2009). Remarkably, MUST is shown to be in decline in both its epistemic and root sense, of which the latter has been affected most strongly (Leech 2003: 234, Smith 2003: 257). This differs from results of the long-term study in Biber (2004b), which finds that epistemic MUST increased and root MUST declined. It is possible that long-term and short-term trends differ, or that the studies' differing underlying assumptions and ways of measuring increases and decreases complicate comparisons.[85] That modal semantics are influential in these developments is widely acknowledged, though: based on spoken 20th century data, Close & Aarts (2010: 165) find that semi-modal HAVE TO is only competing with MUST in root contexts. In epistemic contexts, MUST is also in decline but still by far the preferred option (Close

[84] Brinton et al. (2012) includes a look at deontic modality in Canadian English in a long-term perspective between the 1620s and the present day, confirming the general trends observed on smaller corpora in Dollinger (2006)

[85] Millar (2009: 203–204), working with the TIME Magazine Corpus, also finds epistemic MUST on the rise in proportion to root MUST throughout the 20th century, which goes against the findings in the Brown quartet. As Millar's (2009) work is based on the TIME Magazine Corpus, it has therefore been suggested that his results represent a genre-specific development in journalistic writing (Leech 2011).

& Aarts 2010: 176). HAVE TO has only very recently been emerging in epistemic contexts, and remains quite rare (Krug 2000: 89–90).[86]

Studies of the recent past clearly showcase the influence of the medium: developments in spoken English are markedly different from those in written English. Smith (2003) and Leech & Smith (2006) provide information on spoken British usage in the second half of the 20th century based on two small corpora (80,000 words each, data from 1959-1965 and 1990-1992 respectively). It emerges that the decline of MUST is much steeper in spoken than in written English (Leech 2003: 231). This is in line with findings for earlier centuries, in which genres considered close to the spoken language such as dramatic dialogue were observed to be in the vanguard of change (see 4.2.2). Findings based on the *Diachronic Corpus of Present-Day Spoken English* (DCPSE), consisting of spoken British English between the 1960s and the 1990s, further support these results. Close & Aarts (2010: 176–177) are able to show that already in the 1960s, root HAVE TO was more frequent than root MUST in spoken language. In the 1990s, HAVE TO was approximately three times as frequent as MUST in speech (Close & Aarts 2010: 176–177). These findings lead the authors to re-examine the idea that there could be a potential "link between core modal decline and semi-modal increase" (Close & Aarts 2010: 176–177). As previously discussed, this idea had before been rejected by various scholars based on results from written corpora, which showed an increase in semi-modals but none so great as to make up for the decline in the core modals (see 4.2.1). In spoken language, however, the semi-modals are much more frequent and thus there is no large gap between the fall in frequency of the modals and the rise of the semi-modals. Aarts et al. (2014: 56) confirm these trends and stress that the decline of MUST, for instance, is significantly greater in the spoken than in the written data.

Genre effects can be observed across and within spoken and written material: Johansson (2013),[87] which discusses MUST, HAVE TO, HAVE GOT TO, and NEED TO since the 1990s based on the COCA (American English, ca. 385 million words at the time of

[86] Several other linguistic factors seem to constrain its use as well: it is rare with inanimate subjects and in interrogatives and negation with do-support (Krug 2000: 89–90).

[87] Johansson's (2013) results, especially in terms of frequencies, should not be overestimated because the study is "crude" in some respects, as the author himself points out (374): for instance, root and epistemic senses of the (semi-)modals under investigation were not distinguished, no absolute frequencies are given in the article, etc. Nevertheless, the paper provides some general data for the most recent past, i.e. the developments since the 1990s in American English, and draws attention to some important issues, e.g. genre effects.

the study), makes it clear that different genres exhibit different trends: while HAVE TO is strongly favoured in the 'Spoken' subcorpus, MUST is actually the most frequent option in the 'Academic' subcorpus (Johansson 2013: 375).[88] Different levels of formality across the genres apparently play a role here. It is interesting that Bowie et al. (2013), examining different types of spoken genres in the DCPSE, show that – despite the individual patterns found in different types of spoken genres – MUST is in significant decline in four spoken text categories: formal face-to-face conversation, spontaneous commentary, broadcast discussions and prepared speech (Bowie et al. 2013: 86). This suggests a larger overarching trend.

As in the long-term perspectives, differences between regional varieties emerge. Based on a corpus of 20th century plays, Jankowski (2004: 108) demonstrates that the evolving systems of deontic modality in American and British English show subtly different developments: while changes in both systems include "the loss of *must* as it recedes to the same functions within the respective grammars and obsolesces", there are differences concerning the forms that take over the functions once associated with MUST. American English prefers GOT TO for strong obligation, while it looks like this function is becoming associated with HAVE GOT TO in British English (Jankowski 2004: 106). That the functional specialisation of forms differs between varieties and that there is a "50-year gap between the critical periods of decline", with American English being more innovative, leads Jankowski (2004: 108) to believe that these processes happened independently of each other. Aarts et al. (2014: 56), however, do not exclude the possibility of a "transatlantic influence in the fall in usage of *must*": MUST is declining significantly faster in their written US data than in their written UK data.

Jankowski (2004: 106) further reports that MUST has become "specialized to epistemic modality, performative contexts, frozen expressions and use with stative verbs such as *be* and *have*" in contemporary English. The retreat of MUST to so-called 'performative, 'formulaic', 'relic' or 'rhetorical' uses is also found in other studies. Tagliamonte & Smith (2006: 355), a synchronic study of modality in selected dialects in England, Scotland and Northern Ireland, confirms the increasing limitation of MUST to very specific contexts and formulaic utterances such as *I must say* (Tagliamonte &

[88] The COCA differentiates between the following genres/subcorpora: 'Spoken', 'Fiction', 'Magazine', 'News' and 'Academic'.

Smith 2006: 355). Trousdale's (2003) study of contemporary Tyneside English also mentions that the only non-epistemic contexts in which MUST has any frequency are "lexicalised expressions of the type *I must admit*" (278). In the COCA, Johansson (2013: 377–379) reports, the top 15 verbs following MUST are all speech act verbs (like *say*, *admit*, *warn,* etc.) forming characteristic relic uses. As explained in Myhill (1995: 171), such uses of MUST are "hardly obligations at all", and should instead be considered idiomatic uses.

Concerning the underlying reasons for the observed developments, many studies on 20th-century developments bring forward similar interpretations as the long-term trend studies: Smith (2003: 259), working with 20th-century data, considers MUST "a casualty of a changing society", much like Myhill (1995). Its association with authoritative use of power does not fit in a society where a "democratization of discourse" is taking place and overt power markers are avoided (Fairclough 1992: 201–207). This environment instead favours less authoritative alternatives: HAVE TO, for example, has increased in use considerably since the 1960s (Smith 2003: 249)[89]. One remarkable exception to this overall tendency is found during the Second World War, when "the frequencies of *must* peak at an all time high" in the TIME Magazine Corpus (Millar 2009: 212–213). This temporary restoration of MUST is interpreted as a product of wartime language being less tolerant of ambiguities (Millar 2009: 212–213). Tagliamonte (2004) also associates the decline of MUST in York English with issues of power and societal change: "the obsolescence of *must* may actually be tied to the obsolescence of the appropriate social conditions for its use" (Tagliamonte 2004: 49), namely "contexts in which the speaker has authority over the subject" (Tagliamonte 2004: 51). A different explanation for the general decline of MUST is put forward in Close & Aarts (2010: 177–178): the authors argue that there is a general decline in forms expressing strong commitment,[90] whether that is strong necessity (epistemic MUST) or strong obligation (root MUST). This explanation, it is argued, can also account for the observed decline in epistemic MUST in the 20th century, which is difficult to account for by invoking issues of authority and authoritative language.

[89] Another marker of obligation that profits from this societal development and increases in use is NEED TO. It allows the writer or speaker to "claim that the required action is merely being recommended for the doer's own sake" (Smith 2003: 260). This "strategic use" of NEED to is also reported for the COCA in Johansson (2013: 377–379), along with an increase in a related structure with lexical NEED, *I need you to (do sth.)*.

[90] Close & Aarts (2010: 178) take the term "strength of commitment" from Huddleston & Pullum (2002: 175).

In general, social developments like the ones outlined above seem to be more important to the development of markers of obligation and necessity than purely syntactic considerations, at least for recent history. Smith (2003: 255) shows that the increase in HAVE TO in British English since the 1960s mainly took place in nonsyntactic environments, leaving him to conclude that the availability of HAVE TO where MUST is blocked due to its restricted paradigm was no longer a major factor by then, but had already "peaked in its influence".[91]

Very few studies of MUST and its alternatives include a sociolinguistic perspective, Tagliamonte (2004) and Tagliamonte & Smith (2006) being notable exceptions. Although they are concerned with specific dialectal varieties of British English, they are worth discussing as they systematically include social variables. Age is an important factor, as expected. For expressing epistemic modality in contemporary spoken English in York, MUST is the preferred option (Tagliamonte 2004: 39). In its root[92] sense, however, MUST is obsolescing: only the oldest speakers favour it (Tagliamonte 2004: 49). Younger speakers prefer HAVE TO and HAVE GOT TO. Of note is also the "reversal in the trajectory of change" observed for (HAVE) GOT TO: although both HAVE GOT TO and HAVE TO experienced a period of growth in recent history, HAVE TO is actually the more frequent option among the youngest speakers (Tagliamonte & Smith 2006: 370–371).[93] Young women are identified as the leaders of this change. As MUST is obsolescing, they start favouring HAVE TO over (HAVE) GOT TO, which is socially stigmatised as very colloquial (Tagliamonte & Smith 2006: 373). HAVE TO seems to be stylistically neutral, however (e.g. Krug 2000: 108). At the same time, HAVE GOT TO seems to become specialised for indefinite subjects, especially generic *you*. The results from the COCA for the 1990s onwards fit these observations, too (Johansson 2013: 374–375): MUST is in decline, while HAVE TO is by far the most

[91] Interestingly, (HAVE) GOT TO, a semantically close alternative to HAVE TO, has shown "no signs of growth" in written British English between the 1960s and 1990s (Smith 2003: 259). Most likely, this is the result of a strong 'prestige barrier' for the form with *got*: Krug (2000: 81) goes so far as to call HAVE GOT TO "an item that has received critical remarks in style books throughout its existence", and Leech & Smith (2009: 189) suggest that "the avoidance of forms of *get* in the written language, a well-known taboo, might account for the low and even declining usage of *(have) got to*" in the second half of the 20th century.

[92] Taglimonte (2004: 34-35) uses the term deontic modality instead of root modality but her definition makes clear that she uses deontic in the sense of non-epistemic or root modality.

[93] Note that this is contrary to Krug (2000), who argues that a saturation stage has been reached for both HAVE TO and HAVE GOT TO. According to Krug (2000: 87), HAVE GOT TO is progressively being modalised (*gotta* is used more frequently). Nonetheless, Krug (200: 78) does note a dip in the frequency of HAVE GOT TO in the last 50 years of the ARCHER data (1950-1990). While he interprets this as a result of unsuitable periodisation, it might just as well be an indicator of what Tagliamonte & Smith (2006) describe. Smith (2003: 263) also notes a "stunted development" of (HAVE) GOT TO in the 20th century.

frequently used expression. However, HAVE TO is levelling off. HAVE GOT TO is infrequent throughout the period under investigation, and decreases slightly.

By this point, it should be obvious that the area of English modality is a complex domain involving many forms - some very old (such as MUST), some quite recent (such as HAVE GOT TO). As far as generalisations go, it is safe to say that root MUST is increasingly eschewed in favour of semi-modal alternatives such as HAVE TO and relegated to performative uses such as *I must say*. Epistemic MUST seems to hold its ground. Some structures with MUST, especially *must have* + participle and *must be* are overwhelmingly associated with the epistemic domain, into which alternatives like HAVE TO have made little inroads. It is equally clear that no simple replacement process (of the type 'MUST out, HAVE TO in') is taking place. Furthermore, observed developments are not at all uniform across varieties, registers or groups of speakers. Stylistic associations of some modals also seem to have an impact on the development, e.g. on the rarity of semi-modals in formal text types and the reverse trajectory observed for HAVE GOT TO in the recent past. Spoken language (also reflected in speech-related written texts) seems to be a valuable starting point for any analysis, as it has been made clear that semi-modal alternatives to core modals are "typically colloquial" and "not likely to show up in their true colours in the written language" (Leech 2003: 230). This makes the OBC an exciting resource for the study of this particular phenomenon.

4.2.4 Late Modern grammars on MUST and HAVE TO

For the present investigation, two different – but to some degree interrelated – issues are important: first, to what degree MUST and HAVE TO can be used as alternatives to express obligation, and secondly, whether the form *must* can refer to the past or not.

A survey of Sundby et al. (1991) shows that the first issue, i.e. variation between MUST and HAVE TO, attracts little comment in the 18th century. Only two American grammarians, Webster (1784) and Hutchins (1791), offer remarks: Webster (1784: 100) criticizes the use of *I have to go* as 'improper' and suggests *I must go* as an alternative (Sundby et al. 1991: 212). In a later grammar, the author's policy has changed slightly: Webster (1790: 44) brands *he has got to learn* as improper, but puts forward both *must* and *has to* as acceptable alternatives (Sundby et al. 1991: 330).

Hutchins (1791: 157) comments that HAVE TO is 'vulgar' and that MUST should be used instead (Sundby et al. 1991: 330). In the British grammars, the issue is not dealt with directly but only emerges in an ancillary fashion when Withers (1790) criticizes the use of preposition stranding in the following remark from the *Spectator* 415 (Addison 1712) as 'inelegant': "Our great Modellers of Gardens have their Magazines of Plants to dispose of." As a correction, he suggests *must sell / dispose of their Magazines* but not *have to dispose of / sell their Magazines*, which suggests that he favoured the core modal (Sundby et al. 1991: 426). In general, though, HAVE TO was rarely mentioned.

This relative scarcity of grammatical discussion of HAVE TO continues in the 19th century, even though its frequency was shown to rise. Out of the 16 grammars that I examined, only three acknowledge that HAVE TO can be used to encode obligation or necessity. The earliest grammar in my sample to mention the use of HAVE TO in such a way describes its purpose as follows:

> There is also another mode of expression which, though it does not strictly or positively foretell an action, yet implies a necessity of performing an act, and clearly indicates that it will take place. For example, "**I have to pay a sum of money tomorrow**," that is "I am under a present necessity or obligation to do & future act." (Hiley 1853: 59)

That HAVE TO can be "used to express duty, obligation, necessity, contingency or the coming of some future event" is also noted in Dawnay (1857: 56). Rushton (1869: 204) also acknowledges HAVE TO in his grammar. He also uses HAVE TO in some of the explanatory text of the grammar, e.g. when he speaks of the "Latin language, where the same form has to do double duty [as past and perfect]" (Rushton 1869: 187) or practical rules on the use of WILL and SHALL, which "have to be modified" according to context (Rushton 1869: 195).

Interestingly, some authors who do not explicitly acknowledge HAVE TO nevertheless use it as an expression of necessity or obligation in their texts or in language examples illustrating other points of grammar (Pinnock 1830, Turner 1840, James 1847, Beard 1854, Higginson 1864, Mason 1873). None of the 19th-century grammars contain any critical remarks on the use of HAVE TO.

The second problem, i.e. whether MUST was available in past contexts, and if not, what should be used instead, is not often explicitly discussed in Late Modern

grammars, but only touched upon in passing via the inclusion of MUST in conjugation tables for auxiliary verbs. The reader would often simply find a table with tense forms, in which several cells (such as 'past' or 'perfect') were empty for MUST. Lowth's famous 1762 grammar, for instance, contains a table listing so-called "defective verbs". MUST is listed in the column "present" only; the cells called "past" and "participle" remain empty (83-84). Sundby et al. (1991: 397) only provide one reference to this issue, namely to Priestley (1768: 113) calling MUST an "imperfect" auxiliary, and declaring it "of the present only". Neither Lowth nor Priestley offer any advice on how to deal with this empty cell or recommend alternatives for MUST. To make matters more complicated, not all grammars shared the view that MUST was a present-tense auxiliary only: while considering MUST a "defective" verb, Gardiner (1799: 62) lists it as both a "present" and a "preterite" form in her grammar. Fittingly, Görlach (2001: 123) remarks of the 18th century that "[t]ensing in modals was becoming a problem" which "contemporary grammarians provide[d] insufficient advice" on.

This 'problem' continues into the 19th century, where advice on the issue is also hard to come by: while all 16 British grammars make it clear that MUST is 'invariable' or 'defective', explicit information on the tensing of MUST (and other modals) in the absence of inflection is only found in about half of them, and the recommendations vary between authors. While some acknowledge past MUST, others make it clear that MUST is a present-tense verb only. Overall, nine out of 16 grammars argue that MUST is available as a past form in addition to its use in the present. Allen (1824: 20) says that "[m]ust has no variation on account of number or person" but lists both present and past tense forms for MUST. Bullen & Heycock (1853: 119–120) even go so far to state that "[i]n the present tense, *must* corresponds with *ought* or *it behoves*, and in the imperfect, which seems its true sense, it imports a stronger meaning, as *of necessity*". Other acknowledgments of a past form along with a present form, often only in a verb table but sometimes with further explanation, are found in Hort (1822: 73, 104), Pinnock (1830: 165), Turner (1840: 71), James (1847: 32), Hiley (1853: 38), Dawnay (1857: 57–58) and Mason (1873: 56). These grammars span a broad range of time.

Four grammars clearly state differing opinions. Two of these grammars explicitly argue against past MUST. Crombie (1809), the earliest grammar under

consideration, states that MUST exists only in the present tense (179) and using it as a preterite form is "obsolete" (205). Rushton (1869), one of the later grammars, also points out the unavailability of MUST in the past tense and further remarks that *had to* may be used as a substitute for obligative MUST in this context:

> I have always felt the want of a past tense in this auxiliary [must]. For example, when we wish to translate from German such a phrase as *er musste gehen*, we cannot say 'he must go.' We are obliged to give the sentence a turn: 'he was obliged to go,' 'he was bound to go,' **'he had to go.'** (Rushton 1869: 204)

Two further grammars indirectly argue against past MUST by leaving the cell for 'preterite' or 'past' in the conjugation tables blank (Crane 1843: 240, Curtis 1876: 71).

Independently of that, the more 'modern' alternatives to past MUST, e.g. *had to* for obligative and *must have* + PP for epistemic readings, are mentioned in some grammars or used by the authors in their explanatory texts, showing at least some awareness of variability. Obligative *had to* is thus acknowledged in James (1847: 155), and Hiley (1853: 123). As discussed initially in this section, some other grammarians more generally acknowledged HAVE TO – mostly without specific reference to or use of past *had to* – as an obligative construction. Epistemic MUST *have* + PP appears in Crombie (1809: 179), Hort (1822: 104), Pinnock (1830: 211), Turner (1840: 122), Hiley (1853: 49), Beard (1854: 285) and Higginson (1864: 47).[94]

In general, the discussion of past MUST does not take up much space in the grammars and is clearly not a priority. The most frequent solution, i.e. to simply include a verb table in which MUST is missing in the cells for some of the tenses, both reflects and furthers the contemporary insecurity connected with the proper use of modals that e.g. Görlach (2001: 123) reports.

4.3 Methodological considerations

As previous research clearly illustrates, many factors need to be considered when discussing variation between HAVE TO and MUST in a given context. While it is clear that not everything can be taken into consideration in the present analysis, it is

[94] Reference to *must have* + PP is at times found under names that are foreign to the present-day reader. Hiley (1853: 49) lists *must have had* in a conjugation table for HAVE, calling it 'potential mood, present tense'.

important to make its underlying assumptions explicit. To this end, the present section will justify the choice of variable and variants and discuss the basics of the coding process for this feature.

So far, it has been treated as a given that investigating the domain of obligation and necessity via the modal MUST and the semi-modal HAVE TO is a sensible course of action. Strictly speaking, though, this assumption requires some further comment. After all, variation in the area of obligation and necessity is "not just a two-horse race" (Leech et al. 2009: 98). There are several further (semi-)modals which are closely semantically related to those I focus on, not to speak of alternative expressions of modality such as adjectives (e.g. *necessary*) or adverbs (*surely*) (Huddleston & Pullum 2002: 173). However, there are sound semantic and methodological reasons to focus on MUST and HAVE TO in the present study: after deciding to restrict the investigation to the semantic domain of strong obligation or necessity, the field was narrowed further based on constraints imposed by the available data.

Concerning the first point, this study distinguishes between strong and weak markers of obligation/necessity. Strong obligation markers differ from weaker ones in terms of the consequences in case an obligation is not fulfilled: they are more severe for markers of strong obligation (Bybee et al. 1994: 186). Although the lines drawn by different scholars vary, "*must, have to,* and *got to* are all traditionally characterized as having strong obligation function" (Myhill 1995: 162). Next to MUST, (HAVE) GOT TO and HAVE TO, modal NEED and semi-modal NEED TO are often included in this group (see e.g. Biber et al. 1998: 205, Collins 2009: 33, or Smith 2003: 242) for contemporary English. That this study deals with the Late Modern inventory of modals and semi-modals rather than the contemporary one is not a problem here: the American English data presented in Myhill (1995: 159) shows that the differentiation into weak and strong modals also holds for our period of interest: HAVE TO and GOT TO were rivals to the strong obligation modal MUST, whereas the functions of the weak obligation modal SHOULD were shared by other semi-modal alternatives (*had better* and *ought to*) in Late Modern English.[95]

Practical considerations led to this set of strong obligation markers being narrowed down further. As a meaningful sociolinguistic analysis for gender and class

[95] NEED TO and NEED are not mentioned in Myhill (1995).

requires a large number of tokens, I decided to exclude rare types in the OBC: NEED, NEED TO and (HAVE) GOT TO. A search for all forms of NEED tagged as verbs yields only 420 tokens, which can be assumed to include some false positives. Out of the 420 tokens, semi-modal NEED TO only makes up around 50 occurrences. This is not too surprising, as serious growth of NEED TO has only been identified for the second half of the 20th century (e.g. Smith 2003, Johansson 2013). HAVE GOT TO is also rare: only 53 instances of HAVE GOT with obligative meaning (31 present, 12 past) were found.[96] The earliest instance (25) appears in a trial from 1783:

(25) He said he **had got to** go down to Greenwich or Deptford. (OBC, t17830430-34)

While this instance predates the earliest attestations given in previous work (1860 in Visser 1963-1973: 1479 and 1837 in Krug 2000: 61) by several decades, the scarcity of examples makes it perfectly clear that this option was still a minority choice.[97] Only MUST and HAVE TO are truly frequent in the OBC: they are each attested more than 1,000 times.

It is crucial to ensure that working with these two options does not lead to inadvertently adopting the above-mentioned misconception of the 'two-horse race' in the domain of obligation/necessity. This is why, in addition to acknowledging that the present work can only shed light on the development of the most frequently attested alternatives, this study will take care to provide a context-sensitive exploration of these options, e.g. with regard to semantics (epistemic – root) and temporal reference (past – present). After all, not all constructions[98] involving HAVE TO and MUST are interchangeable in all contexts. While MUST *have* + participle, for instance, is an option in past epistemic contexts, it is not suitable for past root contexts, where other alternatives (like (26)) need to be found (see also 4.1.2):

(26) I **had to** go to a sugar bakers in Wentworth-street, for a hogshead of sugar (OBC, t18131027-2)

[96] GOT TO (without HAVE) is only found once in the OBC, in one of the most recent *Proceedings*: *I found written across my letter in red ink: "Got to be paid."* (OBC, t19100426-40).
[97] The same picture presents itself in the CLMET-drama: only 43 examples of (HAVE) GOT TO are found in the corpus, as opposed to hundreds of instances of MUST and HAVE TO.
[98] For studies with a focus on modal constructions in contemporary English, see Kennedy (2002) and de Haan (2012).

It is clear that we are dealing with a nuanced system. It is essential, then, to only compare options that are true alternatives.

A necessary preparatory step to this end was to exclude those contexts where there is only one option to begin with (see Tagliamonte & Smith 2006, Depraetere & Verhulst 2008, Close & Aarts 2010). The OBC was searched for the forms *must*, *has to*, *have to*, *had to*, potential contracted forms of HAVE TO (*'s to*, *'ve to*) and the variant *hafta*. A good deal of manual post-processing of the search results was necessary to ensure that only instances that a) express modal semantics and b) allow variation between MUST and HAVE TO were considered. To narrow the results down to those with modal meaning, all examples with ambiguous or clearly non-modal readings were excluded. This comprised instances like (27) or (28), in which HAVE TO does not represent a semi-modal of obligation or necessity.

(27) I stopp'd about five Minutes, to hear what they had to say.
(OBC, t17381206-4)

(28) I asked him what right he had to bring it there.
(OBC, t17670218-28)

Another problem is presented by semantically ambiguous structures. Some of these were already excluded by default due to the decision to only search for contiguous HAVE TO. This excludes examples with adverb interpolation and structures involving an object, which often support both a possession and an obligation reading (Dollinger 2006: 292–293).[99] One example including an object is presented in (29).

(29) [H]e used to work for me when I **had** clock-work **to** do.
(OBC, t17450227-12)

In light of this ambiguity, such utterances are best left out of the analysis. More generally, examples where an obligation/necessity reading was in doubt were excluded. Utterances of the type exemplified in (30), where the (semi-)modal is followed by ellipsis, were also discarded:

(30) Yes, I must. / It had to.

[99] Examples involving adverb interpolation are even thought to represent an intermediate step in the grammaticalisation of HAVE TO (see also Brinton 1991), which makes them "categorically ambivalent" (Dollinger 2006: 292–293).

They provide insufficient information on the temporal reference and the semantics of the (semi-)modals (also see Close & Aarts 2010: 165).

To concentrate only on contexts in which both options are true alternatives, the dataset was further reduced: negated examples, formulaic expressions and syntactic uses of HAVE TO were removed. All negated examples were discarded because "the scope of negation is different for HAVE TO (absence of necessity, 'not necessarily') and MUST (prohibition, 'necessarily not')" (Depraetere & Verhulst 2008: 16), as examples (31) and (32) illustrate:

(31) You **must not kill** my cousin. (OBC, t18680106-142)

(32) You don't have to kill my cousin.

In (31), an original example from the OBC, the modal is outside the scope of the negation and the sentence can therefore be paraphrased as 'it is necessary for you to not kill my cousin'. In example (32), the modal is inside the scope of the negation and the sentence may be paraphrased as 'it is not necessary for you to kill my cousin'.[100] The formulaic expressions *must needs* and *needs must*, rarely represented to begin with (12 and 1 instance(s), respectively), were removed because there is no parallel structure with HAVE TO.

Finally, all syntactic uses of HAVE TO, i.e. instances in which MUST cannot serve as an alternative due to its incomplete verbal paradigm, were excluded. This concerned non-finite forms[101] of HAVE TO (33) and combinations of HAVE TO with other modals (34).

(33) I was not quite satisfied **to have to** wait till the following Friday week for the shares.
(OBC, t18920307-312)

(34) I then cautioned him that anything he might say I **might have to** give in evidence against him.
(OBC, t18910209-205)

[100] Close & Aarts (2010: 171–172) also exclude negated forms in their study on MUST, HAVE TO and HAVE GOT TO based on the differing scopes of negation for these variants.

[101] The *-ing* form is, of course, also a non-finite form. However, the search terms automatically excluded examples like the following one: *I was angry at **having to** give my name and address* (OBC, t18991120-33). 22 instances of *having to* with modal meaning are found in the OBC.

Note that, based on my remarks in 4.1.2, I do not consider past tense settings a syntactic environment in this sense in the Late Modern period (also see Dollinger's (2006) assessment of past MUST as nonsyntactic in LModE).

The remaining instances of MUST and HAVE TO in both corpora were coded for the factors and levels shown in Table 12. Factors are in capital letters.

Factor	**Levels**
VERB[102]	MUST HAVE TO
TYPE OF MODALITY	root epistemic performative ambiguous
TIME REFERENCE	past non-past
PERIOD	1720-1769 1770-1819 1820-1869 1870-1913

Table 12. Coding for analysis of MUST/HAVE TO

In addition to this, tokens in the OBC were automatically coded for the social parameters of speakers (GENDER and SOCIAL CLASS).

Coding for semantic subcategories (labelled TYPE OF MODALITY in Table 12) is crucial for reasonable comparisons, as the semantic subtype has a bearing on which alternants exist for a modal expression (Aarts et al. 2013: 20). In addition to the well-known categories 'root' and 'epistemic' already described (4.1.1), I also recognise ambiguous cases and performative uses, both of which require some comment. I used the label 'performative' for those expressions that are best understood as idiomatic uses of HAVE TO or MUST with very little or no obligative component (see 4.2.3). The label 'performative' is fitting because they "occur where the speaker is carrying out the action denoted by the verb" (Close & Aarts 2010: 174), as illustrated in (35) and (36).

(35) I **have to inform** you that at our first meeting I deceived you.
(OBC, t18081130-11)

[102] The dependent variable is shaded in grey in this table, as in all tables on coding.

(36) I told Mr. Jones the bill was my property, and that **I must insist** upon knowing how he came by it.
(OBC, t17500530-23)

Despite a thorough examination of the context in which a semi-modal or modal occurs, it was not always possible to establish what the intended meaning of the utterance was or to make an argument that would clearly favour one interpretation over another. Example (37) is one such case:

(37) Q. When a man has said you swear so, and you say no, I did not, that is wrong, **you must recollect whether you ever said so or not**. Did not you twice say so to the clerk, and twice correct the clerk? - Not to my knowledge.
(OBC, t17940219-74)

It is unclear to me whether *you must recollect* carries obligative meaning ('you are obliged to recollect what happened because your testimony is important') or epistemic meaning ('it can be logically inferred that you are able to recollect this because you have already testified on other details of the matter in this trial'). I classified such cases as 'ambiguous'.

When coding for TIME REFERENCE, I distinguish between 'past' and 'non-past.' Note, however, that I do not use the term 'tense' here, because this category is not about the tense of the modal or semi-modal as such (it could be argued that modals do not express any tense information at all), but concerns the difference between modals occurring in past contexts and in present contexts. It may be objected that the central modals developed "purely modal, non-past use of the preterite forms *would, should, might, must*" (Rissanen 1999: 235) and that including past contexts therefore makes little sense. For present-day English, I agree with this assessment. However, in the Late Modern period, this development is not yet completed. As explained, MUST had not yet entirely lost its past-referring use, which warrants a distinction between past and present MUST. This procedure also makes sense in light of the material in the OBC: after all, speakers in a trial setting use much of their discourse to elaborate on what happened in past situations. For HAVE TO, verb forms are enough to differentiate between past and present reference. In the case of MUST, all examples had to be checked individually based on context. As mentioned already, oblique uses were also coded as past-referring. Where I could not establish the temporal setting for an instance of MUST, I included it in neither the discussion of past or non-past MUST.

4.4 Findings and discussion

This section summarizes the analysis of MUST and HAVE TO in the OBC and in the CLEMT-drama and provides an interpretation of the findings. First, a general overview of the diachronic development of the variants in the OBC is presented, including contemporary comments on the usage of these forms (4.4.1). After that, root and epistemic contexts are discussed separately (in 4.4.2 and 4.4.3, respectively), before conclusions are presented in 4.4.

4.4.1 MUST and HAVE TO in Late Modern English: general trends

The OBC yields 5,787 relevant instances of MUST and 1,282 examples of nonsyntactic HAVE TO, which are shown in Table 13 according to semantic types:

	HAVE TO	MUST	Total
root	1,240	2,916	4,156
epistemic	18	2,686	2,704
performative	2	101	103
ambiguous	21	84	105
total	**1,281**	**5,787**	**7,086**

Table 13. (Semi-)modals by type of modality in the OBC

These numbers confirm that type of modality greatly impacts the distribution of MUST and HAVE TO. Among those categories well represented in the corpus, i.e. root and epistemic modality, it is only root modality that shows substantial variation. Epistemic modality is overwhelmingly encoded with MUST, just like the much rarer performative modality. As it is only a marginal type in this corpus, performative examples will not be studied further. Ambiguous cases in which the type of modality could not be clearly identified are also excluded from further discussion. This leaves 6,860 examples (shaded in Table 13), i.e. 2,704 with epistemic meaning and 4,156 with root meaning, for analysis.

The diachronic development of MUST and HAVE TO in the OBC for both epistemic and root meaning is shown in Table 14.

Modality	Epistemic			Root		
	MUST	HAVE TO	% of HAVE TO	MUST	HAVE TO	% of HAVE TO
1720-1769	387	0	0.0%	870	14	1.6%
1770-1819	788	0	0.0%	856	73	7.9%
1820-1869	820	4	0.5%	714	375	34.4%
1870-1913	691	14	2.0%	476	778	62.0%
Total	2,686	18	0.7%	2,916	1,240	29.8%

Table 14. Nonsyntactic MUST and HAVE TO in the OBC, by period and type of modality

It is clear that MUST is favoured in all periods to express epistemic necessity. HAVE TO represents a marginal choice. Only in 18 out of 2,768 examples, i.e. less than 1% of the time, HAVE TO was chosen to encode epistemic necessity. For root contexts, however, the situation is very different: HAVE TO is rapidly gaining ground throughout the Late Modern period. In the final period under investigation, HAVE TO has become the preferred choice in root contexts, claiming more than 60% of the share.

The relative scarcity of HAVE TO in the 18th century OBC texts is compatible with what we see in the grammars in the 18th century, where it is almost never discussed. To be a target for discussion, a variant in question would probably need to be more frequent. That the comments we do find tend to be rather critical of the use of HAVE TO supports the assumption that HAVE TO was an unconscious innovation that originated in spoken registers. Against the long-standing variant MUST, it must have seemed 'informal' to at least some commentators.

4.4.2 Root meaning

The OBC contains 1,240 instances of HAVE TO and 2,916 instances of MUST with root meaning in nonsyntactic contexts. Following the model selection procedure outlined in 3.5, a logistic regression model including the predictors PERIOD (in which an utterance was produced), SOCIAL CLASS and TIME REFERENCE (of the verb) was fitted. Its results are outlined in Table 15. The estimates in the second column indicate the log odds of MUST being used.

	estimate *b*	SE	z value	p-value	confidence intervals	
					2.5%	97.5%
Intercept	5.0637	0.3426	14.780	<0.001	4.441954	5.7979101
TIME REFE-RENCE=past	-1.5329	0.1100	-13.936	<0.001	-1.750822	-1.3194671
CLASS= lower	-0.6389	0.1121	-5.701	<0.001	-0.859528	-0.4200060
PERIOD= 1770-1819	-1.1437	0.3607	-3.171	<0.01	-1.905354	-0.4764931
PERIOD= 1820-1869	-3.1257	0.3329	-9.390	<0.001	-3.843385	-2.5249355
PERIOD= 1870-1913	-4.4172	0.3343	-13.215	<0.001	-5.137409	-3.8137427
Concordance Index *C*		0.85				

Table 15. Output of logistic regression including predictors TIME REFERENCE, CLASS and PERIOD; based on OBC

In essence, the regression predicts that – all other things being equal – the chance of MUST being used is higher in present contexts than in past contexts (see the minus sign in the column called 'estimates' for TIME REFERENCE = past), higher among higher-class speakers than lower-class speakers and higher at the beginning of the period under investigation than at the end (note the progressively worsening odds for MUST in the last three rows of the table). For purposes of illustration, these effects will be discussed individually in the following.

As an ongoing change is mapped here, it makes sense that time as a factor is significant. The model thus predicts that the share of MUST diminishes steadily: the effect plot in Figure 9 illustrates this nicely.

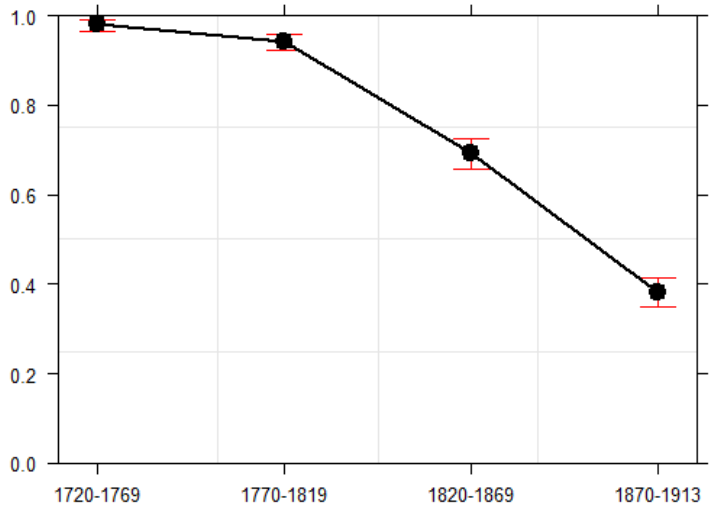
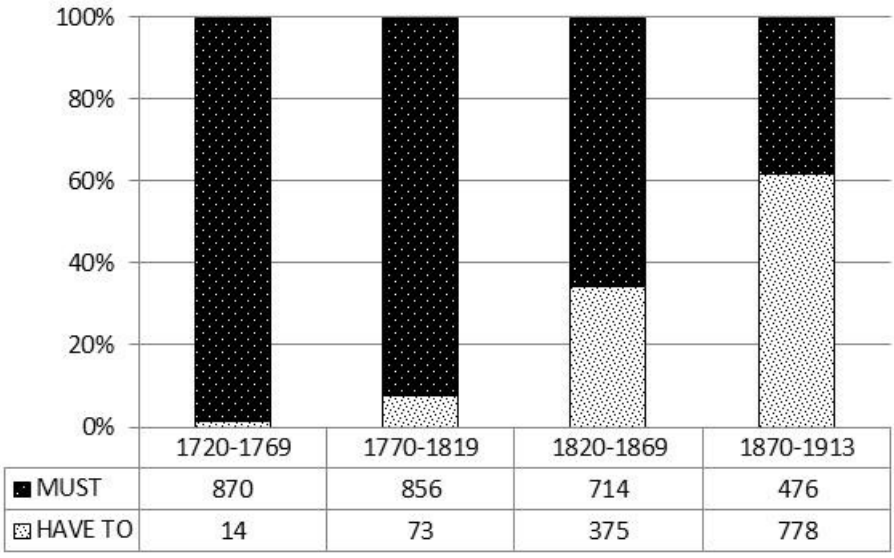


Figure 9. Effect of factor PERIOD on likelihood of MUST (data: OBC)

The likelihood of MUST as the variant of choice to express obligation is very high indeed in the early Late Modern period: the model predicts 98.1% in the period spanning 1720 to 1769, and 94.2% in the period between 1770 and 1819. In 1820-1869, the likelihood of MUST drops to 69.3%, meaning that HAVE TO is predicted for almost a third of all expressions of obligation. In the final period under investigation, HAVE TO is predicted to be more likely than MUST for the first time: MUST only reaches a predicted likelihood of 38.2%. The actual development in the corpus, on which these calculations are based, is shown in Figure 10.



	1720-1769	1770-1819	1820-1869	1870-1913
MUST	870	856	714	476
HAVE TO	14	73	375	778

Figure 10. MUST and HAVE TO (nonsyntactic, root meaning) in the OBC, by period (N = 4,156)

The proportion of HAVE TO, while extremely small in the beginning (14 out of 884 tokens, i.e. roughly 1.6%, in the period 1720-1769), is clearly growing, and finally reaches more than 60% of the overall share in the final period, 1870-1913.

In terms of social factors, the regression identifies a class effect. Figure 11 summarizes the model's predictions. It shows the predicted likelihood of MUST being used in the groups of higher-class speakers and lower-class speakers in the OBC.

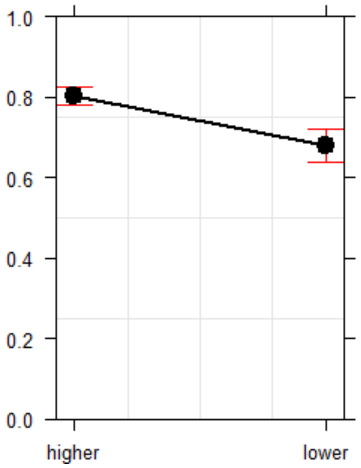


Figure 11. Effect of factor CLASS on likelihood of MUST (data: OBC)

The confidence intervals do not overlap, which indicates a significant difference between these groups: indeed, the model predicts a significantly smaller likelihood of MUST among lower-class speakers, namely 68.3% as opposed to 80.4% among higher-class speakers. Based on the observations extracted from the corpus, the diachronic development of all observed instances in the OBC is shown in Figure 12:

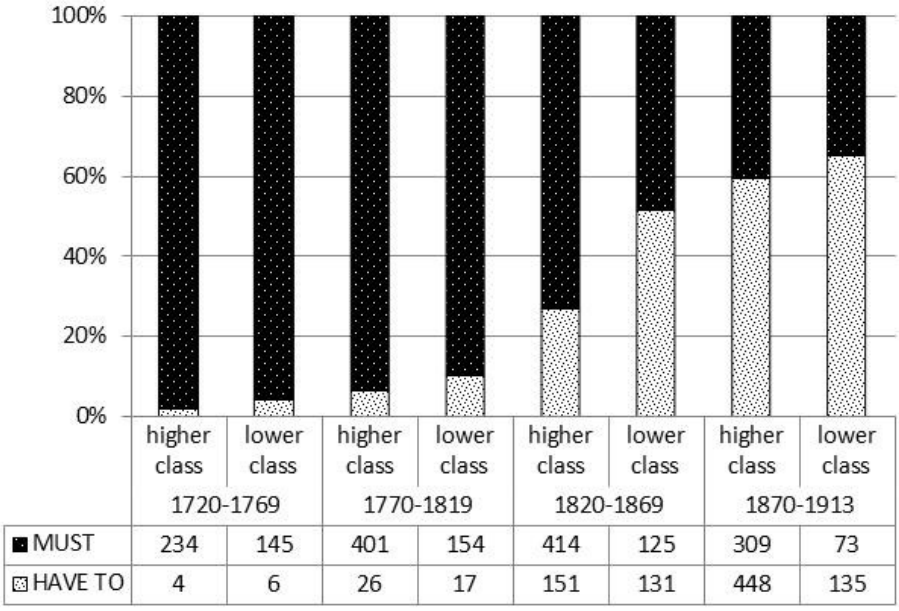


Figure 12. MUST and HAVE TO in the OBC, by class and period (N =2,773)

In each period, the lower-class speakers use proportionately more of the new variant than the higher-class speakers. This is in keeping with HAVE TO being a subconscious innovation that originated in spoken conversation. It is conceivable that people with

limited access to education would be more accepting of so-called 'informal' variants being used in settings like a courtroom than higher-class speakers, who may have been more concerned with linguistic matters and hesitant to introduce or adopt new forms that might challenge established usage.

A linguistic variable was also identified as significant for the distribution of HAVE TO and MUST: the time reference of the utterance. Utterances with past reference more readily accept HAVE TO, as the effect plot in Figure 13 indicates. The regression predicts a rather sizeable difference between past and present contexts: the likelihood of MUST in present contexts is 89.8%, compared to only 65.5% in past contexts.

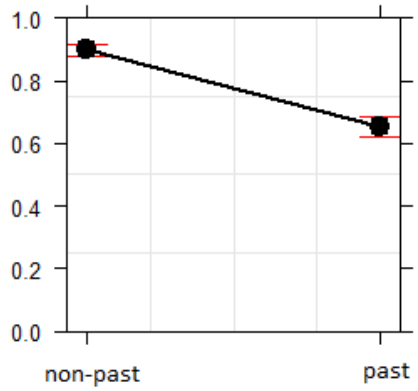


Figure 13. Effect of factor TIME REFERENCE on likelihood of MUST (data: OBC)

To add diachronic information to the picture, Figure 14 shows the distribution of HAVE TO and MUST, by time reference, across four periods in the OBC.

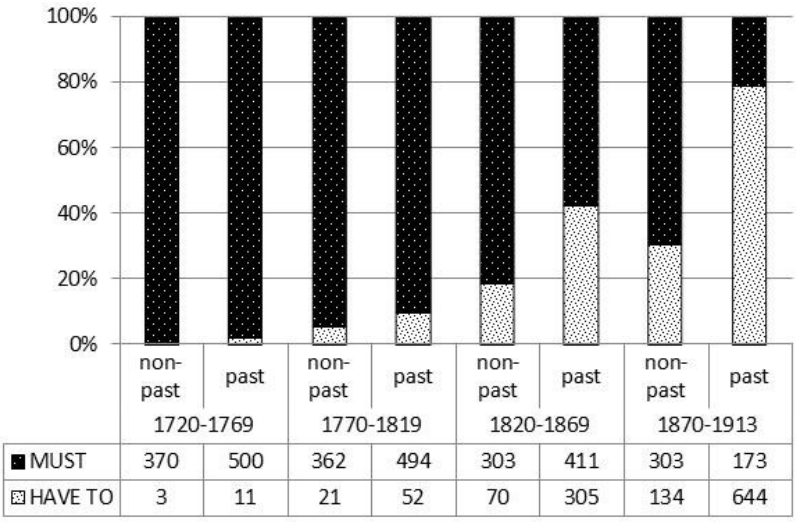


Figure 14. MUST and HAVE TO in the OBC, by time reference and period (N = 4,156)

It seems that HAVE TO was most readily adopted in past contexts, perhaps because MUST was beginning to be seen as unsuitable in such environments (see 4.4.1 on uncertainties concerning modals and tensing in the 18th century).

Finally, some comment is required on the social factors not discussed so far, gender and role. Gender, while frequently a significant variable in sociolinguistic work, turned out to have no significant impact in the present case and was thus dropped from the regression model. Neither women nor men can be said to lag or lead in the change from MUST towards HAVE TO as the preferred marker of root obligation While the factor role had to be excluded from the regression model for methodological reasons (see 3.5), it is worth looking at in isolation: Table 16 summarises the distribution of HAVE TO and MUST across various roles. It shows that the use of expressions of obligation varies between groups in terms of frequency and in terms of preferred variants.

Role	HAVE TO	MUST	expressions of obligation (HAVE TO + MUST)	words per group in the OBC	expressions of obligation p100tw[103]
defendant	108	0	108	817,235	13.2
judge	3	42	45	437,414	10.3
lawyer	15	42	57	766,273	7.4
victim	217	1,459	1,636	2,315,587	70.7
witness	857	1,457	2,270	7,705,781	29.5

Table 16. Root MUST and HAVE TO in the OBC, by role (N = 4,200)

It is striking that defendants do not use MUST at all – although this variant is preferred among all other roles in the corpus. As MUST is often characterised as a very authoritative variant (Myhill 1995), it may be argued that defendants are reluctant to use it in the courtroom. They are in a weak position to begin with and perhaps do not want to be seen as antagonistic. However, a look at the list of results shows that most uses of MUST and HAVE TO are not part of expressions of obligation aimed at others in the courtroom but part of retellings of the alleged crime, and thus often retellings of the words of others.[104] This does not encourage conclusions based on the alleged 'authoritative' nature of individual variants.

[103] The abbreviation 'p100tw' is used for 'per 100,000 words' in the present study.
[104] That trial participants routinely repeat other people's (alleged) utterances is an important issue for Widlitzki & Huber's (2016) study of swearing and taboo language in the OBC. Swearing aimed at others in the courtroom

The rightmost column of Table 16 (i.e. expressions of obligation p100tw) further highlights an important peculiarity of courtroom data: as mentioned in 3.2.3, a speaker's role constrains their types of interactions and thus also the likelihood of certain phenomena being found in their speech. Victims and witnesses show the highest rates of expressions of obligation in the proceedings: 70.7 and 29.5 instances per 100,000 words, respectively. These are the groups that contribute the bulk of testimony, while records of defendants' speech are often rather limited for procedural and political reasons. Expressions of obligation are mainly found in retellings of prior discourse. This type of talk is largely missing in lawyers' and judges' speech, which explains the low incidence of such expressions for these roles.

To contextualise the findings from the OBC, a comparison with the CLMET-drama, representing another speech-related text type, is the next analytical step. Table 17 displays all relevant instances of MUST and HAVE TO in the drama corpus and the proportion of HAVE TO by period.

	HAVE TO	MUST	% of HAVE TO
1710-1780	3	665	0.4%
1780-1850	4	454	0.9%
1850-1920	145	469	23.6%

Table 17. MUST and HAVE TO in the CLMET-drama, by period

It is immediately apparent that HAVE TO is off to a very slow start in the drama corpus. For a meaningful comparison to the OBC, a new regression analysis was run, including all nonsyntactic root uses of MUST and HAVE TO in the OBC and the CLMET-drama.

The following factors were found to significantly impact the choice between HAVE TO and MUST: PERIOD (this time operationalised as the three pre-set CLMET periods), TIME REFERENCE and CORPUS. The results of a logistic regression including these factors are displayed in Table 18. The estimates in the second column refer to the likelihood of MUST.

generally does not take place. Instead, swearwords and taboo language are encountered in reports of earlier spoken interaction. Summarising, one can say that speakers in the OBC do not use 'bad language' themselves, but report its (alleged) use by others.

	estimate *b*	SE	z value	p-value	confidence intervals	
					2.5%	97.5%
Intercept	6.0064	0.2365	25.397	<0.001	5.5640009	6.4944530
PERIOD= 1780-1850	-2.1995	0.2243	-9.805	<0.001	-2.6650730	-1.7817522
PERIOD= 1850-1920	-4.4473	0.2183	-20.374	<0.001	-4.9027298	-4.0431974
CORPUS= OBC	-0.6044	0.1173	-5.152	<0.001	-0.8355079	-0.3753923
TIMEREF= past	-1.9802	-0.0930	-21.292	<0.001	-2.1641951	-1.7995421
Concordance Index *C*		0.88				

Table 18. Output of logistic regression including predictors PERIOD and CORPUS; based on OBC and CLMET-drama

Again, TIME REFERENCE and PERIOD play a role for the choice between MUST and HAVE TO. Past-tense contexts lower the chances of MUST occurring, just as more recent periods do. The effect plots (Figure 15 and Figure 16) illustrate this:

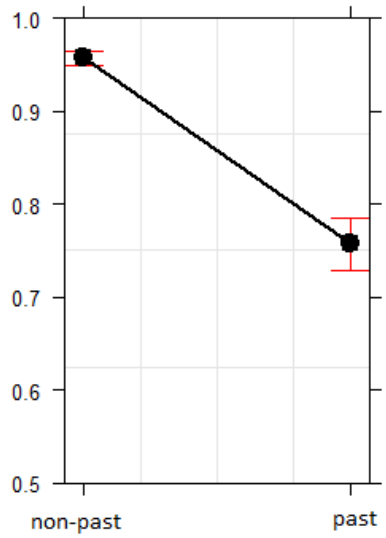


Figure 15. Effect of factor TIME REFERENCE on likelihood of MUST (data: OBC and CLMET)

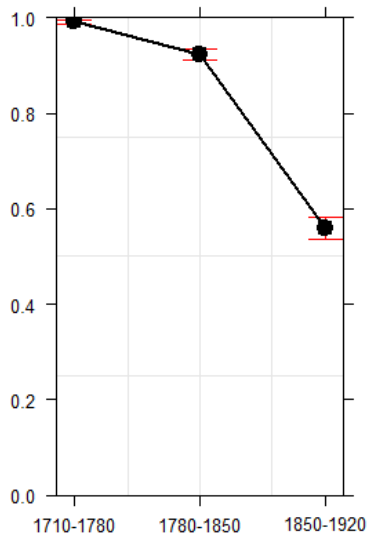
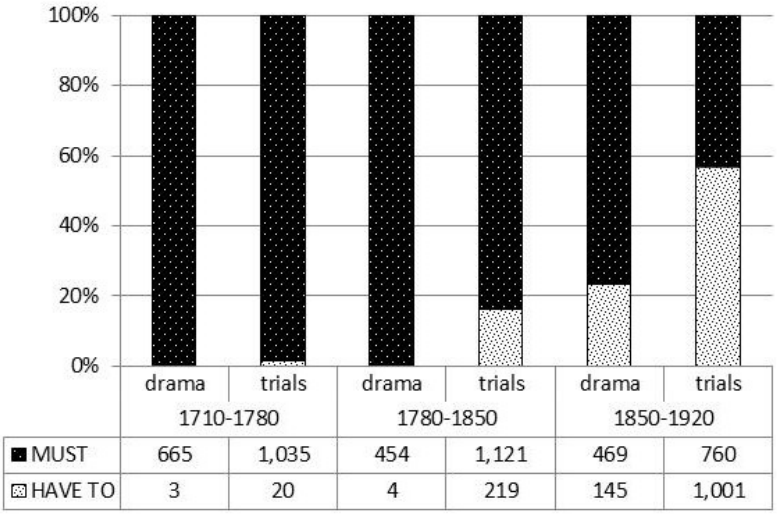


Figure 16. Effect of factor PERIOD on likelihood of MUST (data: OBC and CLMET)

The calculated likelihood of MUST differs significantly by time reference (past: 75.7%, non-past: 95.8%) and period (1710-1780: 99.1%, 1780-1850: 92.3%, 1850-1920: 55.9%). In comparison with the model run only on the OBC, the likelihood of MUST is a little higher in the model with both corpora. This already indicates that the CLMET-drama is more conservative and favours MUST more strongly.

The effect of the corpus, shown in line 4 of Table 18, confirms the significant effect of the different genres/corpora. The effect as calculated by the model is the following: the predicted likelihood of MUST for the OBC is 88.5%, that for the CLMET 93.3%. Figure 17, which provides a comparison between the actual distributions of MUST and HAVE TO in the two corpora, also attests to this difference:



	drama	trials	drama	trials	drama	trials
	1710-1780		1780-1850		1850-1920	
MUST	665	1,035	454	1,121	469	760
HAVE TO	3	20	4	219	145	1,001

Figure 17. MUST and HAVE TO in the CLMET (drama) and the OBC (trial), by period (N = 5,896)

Trial proceedings (the OBC) are less favourable to MUST – and thus more favourable to the newer variant HAVE TO – than the dramatic texts.

This development is interesting in light of what Krug (2000: 196) proposed about innovation diffusion and distance to the spoken language (see 2.2.2). According to his S-curves for different genres, we would expect to find the drama corpus in the lead, as drama is closest to the 'informal spoken' end of the continuum he proposes. It would be expected to adopt spoken innovations fastest. Trials, which would feature among spoken formal language in his model, would lag behind. However, the OBC trials are ahead of the plays in each period, accepting the informal variant HAVE TO more quickly. This may indicate that the material in the OBC is closer to the spoken end of the continuum than Late Modern drama and that the model underestimates the ability of trial proceedings to integrate spoken features. The variable in question may facilitate the observed development: HAVE TO, while taken over from informal language, was not really stigmatised and thus not a high-risk choice even in a formal

setting – contrary to features like *I says* (see 6.1.3), which blatantly disregard standard concord and were commented on. For HAVE TO, even more formal speech such as courtroom discourse seems a welcoming environment.

4.4.3 Epistemic meaning

As noted in 4.4.1, there is practically no variation between MUST and HAVE TO in epistemic contexts (18 observations of HAVE TO vs. 2,686 of MUST; see Table 14). MUST is clearly the preferred choice.[105] The rare examples of HAVE TO, such as (38), only appear in the later periods:

(38) You were behind me, and the blow **had to** come over my shoulder to strike my wife, because I had her in my arms […]. (OBC, t-18690816-785)

This is not surprising, as epistemic uses of HAVE TO remain rare even in present-day English (see e.g. Coates 1983: 57, Collins 2009: 59). We would thus not expect to find many in earlier stages of the development. The situation in the drama corpus is even more extreme than in the trials. In the CLMET-drama, we find 519 examples that express epistemic necessity, distributed relatively evenly across the three time periods (1710-1780: 205, 1780-1850: 155, 1850-1920: 159). Without exception, MUST is used.

While a variationist analysis of MUST and HAVE TO therefore clearly makes no sense, there are other issues of interest that can be investigated using the OBC data, such as the types of constructions that MUST occurs in. Especially the pattern MUST + *have* + participle (*she* ***must have*** *gone home*) is said to increase during the Late Modern period (see e.g. Biber 2004b, Furmaniak 2011; also 4.1 and 4.2). In fact, this pattern and its solid connection with epistemic meaning is usually cited as one of the reasons why MUST is not obsolescing. To see whether the OBC data also confirm this, Table 19 shows a breakdown of all epistemic uses of MUST by different modal verb phrase structures, according to the typology used in Kennedy (2002).[106]

[105] If all the (previously excluded) syntactic uses of HAVE TO with epistemic meaning are added to the count, the figures do not change much: 82 instances of HAVE TO (instead of 18) stand against 2,686 instances of MUST.

[106] The name 'modal verb phrase structures' is taken from Kennedy (2002). He distinguishes 9 phrase structures; the numbers in the table follow his categorisation. Some structures do not appear in the OBC data, which is why not all numbers from 1-9 are found in the table. Aarts et al. (2014) also distinguish different modal verb phrase structures, or 'modal (verb phrase) patterns', in their terminology. They largely use the same basic categories as Kennedy (2002), but also further distinguish subcategories.

	1720-1769	1770-1819	1820-1869	1870-1913	Total
(2) MUST + infinitive	242	369	120	93	824
(3) MUST + *be* + past participle	28	25	15	6	74
(4) MUST + *be* + present participle	0	0	0	3	3
(5) MUST + *have* + past participle	98	336	579	463	1,476
(7) MUST + *have* + *been* + past participle	18	57	100	113	288
(8) MUST + *have* + *been* + present participle	1	1	6	13	21

Table 19. Epistemic MUST in different modal verb phrase structures in the OBC, by period

The figures show that most examples of epistemic MUST occur in two frequent patterns, namely 2 and 5 (shaded cells), exemplified in these examples:

(39) MUST + infinitive (structure 2):
That **must be** a mistake. (OBC, t18901215-122)

(40) MUST + *have* + past participle (structure 5):
All right; if I had any bad money somebody **must have put** it in my pocket […]. (OBC, t18990912-591)

Their developments are compared in Figure 18:

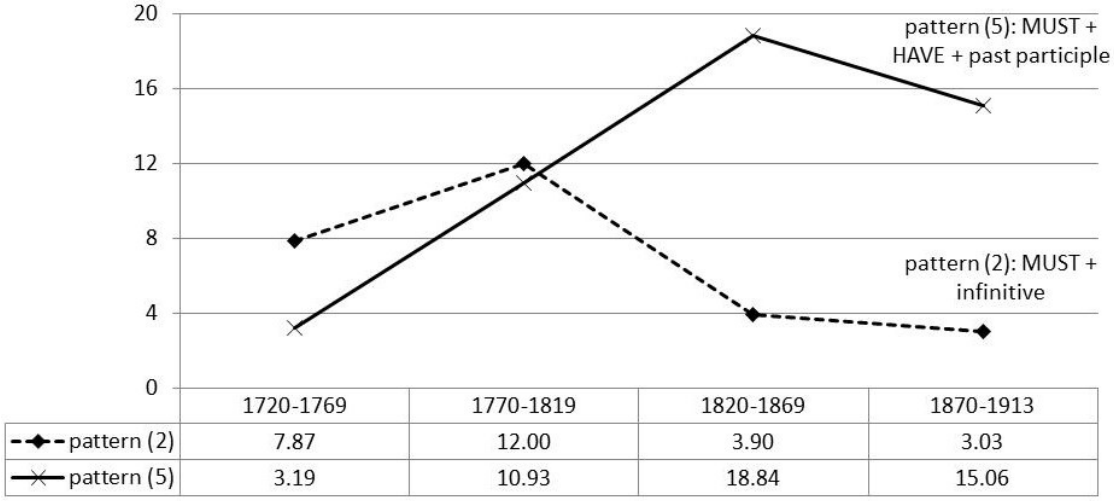


Figure 18. Epistemic MUST in two highly frequent patterns in the OBC (pmw)

Verb phrase structure 2 is clearly increasing, as we would expect based on previous work. After a first occurrence in the OBC in 1732 (OBC, t17320114-12: *he **must have had** some money*), its incidence per million words increases almost sixfold between the first and third period under consideration, from 3.19 to 18.84. The dip in the fourth period could be a sign that the construction is levelling off. For the pattern MUST + infinitive, however, a noticeable drop in use is recorded.

Data from the CLMET-drama show the same general trends. The absolute frequencies for MUST in various phrase structures are shown in Table 20.

	1710-1780	1780-1850	1850-1920	Total
(2) MUST + infinitive	163	117	101	381
(3) MUST + *be* + past participle	8	7	0	15
(4) MUST + *be* + present participle	0	1	3	4
(5) MUST + *have* + past participle	28	29	50	107
(7) MUST + *have* + *been* + past participle	5	1	2	8
(8) MUST + *have* + *been* + present participle	1	0	3	4

Table 20. Epistemic MUST in different modal verb phrase structures in the CLMET-drama, by period

Once again, the constructions that make up the most epistemic uses of MUST are numbers 2 and 5. A look at relative frequencies shows the same patterns as in the OBC (Table 21).

	1710-1780	1780-1850	1850-1920
(2) MUST + infinitive	39.96	28.68	24.76
(5) MUST + *have* + past participle	6.86	7.11	12.26

Table 21. Frequencies per 100,000 words for modal verb phrase structures with MUST in the CLMET-drama

MUST + infinitive is declining and MUST + *have* + past participle is growing. Partly, these developments are interrelated, it seems: note that the category MUST + infinitive includes patterns where MUST occurs with past and with present time reference. With the increasing use of MUST + *have* to signal past tense for epistemic MUST (culminating

in it becoming obligatory in that context during the Late Modern period; see Molencki 2003: 85), MUST + infinitive is relegated to present tense uses only, which explains why this construction becomes less frequent in general. The variation between MUST *have* + past participle and past MUST is shown in Figure 19.

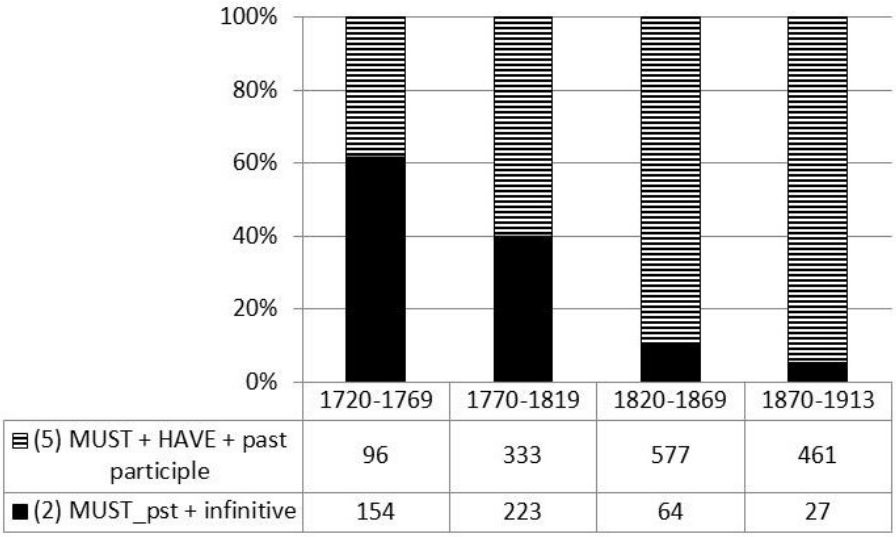


Figure 19. Epistemic MUST in the OBC: past MUST + infinitive and MUST *have* + past participle (N =1,935; $\chi 2 = 442.01$, df = 8, $p < 0.001$)

In the first period under investigation, MUST as a past form is still preferred (with over 60% of the share), but by the third period, MUST *have* + past participle clearly has become the standard option. MUST with past reference is used less than 10% of the time in the latter two periods. A significant decline takes place.

4.4.4 Conclusions

This overview has provided some first insights regarding the social dimension of morphosyntactic variation in LModE, the impact of the linguistic context for the choice between HAVE TO and MUST, and the methodological benefits and challenges of studying this phenomenon in the OBC.

After identifying root meaning as the context in which variation and change between MUST and HAVE TO can be observed, the analysis showed that HAVE TO becomes steadily more frequent throughout the Late Modern period, turning from a minority option (with a proportion of less than 2%) to the preferred choice of speakers

(used more than 60% of the time). As the analysis deliberately excludes all contexts in which HAVE TO would be the only option due to the paradigmatic limitations of MUST, this rise cannot be interpreted as a consequence of HAVE TO simply acting as a formally more versatile 'substitute'. It was shown, though, that the proportion of HAVE TO was greater in past contexts than in present contexts. In the final period under investigation in the OBC, it was in this context that HAVE TO is used in almost 80% of cases. MUST with past time reference is obsolescing, and only retained in some backshifting contexts. In all periods, though, there is variation between MUST and HAVE TO in both past and present contexts. This supports the previous decision to treat past contexts as nonsyntactic.

Class was identified as a significant social factor for the feature. Lower-class speakers use the incoming variant HAVE TO more than higher-class speakers. If change in the system of obligation and necessity indeed originated in spoken language and in informal conversation, then this makes sense. Lower-class speakers in the OBC may be less hesitant to use the more recent and more conversational feature in the courtroom. Higher-class speakers were likely more influenced – at least from the mid-18th century onwards – by the ideas of formality, politeness and correctness in language and thus may have been more resistant to innovation. After all, though, the choice between MUST and HAVE TO is, in terms of linguistic scrutiny, a rather low-risk, perhaps even an 'invisible', choice. HAVE TO can hardly be considered a stigmatised feature, and contemporary grammars offered little comment.

As far as epistemic modality is concerned, HAVE TO is no true alternative to MUST, which is clearly the dominant option. While root MUST is in decline, the epistemic use is holding its ground (also see previous work by e.g. Coates 1983, Myhill 1995, Tagliamonte 2004, Collins 2009, Schulz 2011). This does not mean that the situation of epistemic modality is static, though: a rise is shown for the construction MUST + perfect infinitive, and a concurrent fall in the use of MUST + infinitive. It seems that especially MUST as a past form in connection with the infinitive is retreating to more and more restricted contexts (backshifting contexts), and MUST *have* + past participle taking its place. This also shows that these developments can only be meaningfully explained when their connections are taken into account.

Methodologically, some interesting aspects have arisen. Concerning the corpora, we can state that the results from the OBC and the CLMET-drama support each other: the same general trends are observable, expanding on but remaining in line with earlier research on the development of the modals of obligation and necessity. The development from MUST to HAVE TO as the preferred choice to encode obligation progresses at differing speeds in the corpora, though, with the drama corpus lagging behind the OBC. Assuming that the innovations are speech-based, this means that the OBC mirrors speech more closely – at least for this feature. This is supported by the fact that the first instance of the even more innovative variant (HAVE) GOT TO in the OBC (1783) also predates the one from the CLMET-drama in the period 1850-1920. Playwrights may have been more conservative than the spoken conversation at the time.

Finally, the chapter has stressed the importance of identifying contexts in which variation is truly possible and to take into account factors like time reference of the verb or modal semantics. If figures for MUST are presented independently of constructions (e.g. including MUST HAVE), temporal context (e.g. including both past and present MUST), modal semantics (e.g. both root and epistemic examples), this leads to an inaccurate picture and perhaps even to the conclusion that nothing much happened in the domain of modality, while 'under the surface', actually much has shifted around, as this chapter illustrates.

4.5 Summary

The present chapter started with an overview of modal semantics and the history of MUST and HAVE TO (4.1). Subsequently, previous research was reviewed both for the Late Modern period as well as for the more recent past (4.2). After explaining methodological assumptions and basic principles in (4.3), including e.g. the importance of accounting for semantic factors and the issue of time reference in the discussion, the analysis of variation and change in the OBC and in the CLMET was presented in 4.4.4.

The analysis has put me in a position to address the hypotheses formulated in 3.6: A change was expected from MUST to HAVE TO. This can be confirmed for root

obligation; in epistemic contexts, MUST is the dominant option throughout the period. The hypothesis that developments in the OBC and the CLMET-drama will be different can be confirmed in so far that the OBC is ahead of the CLMET as far as this change is concerned. In general, the same trends are visible in both corpora, just with a time lag for the CLMET. As neither of the variants can count as 'heavily stigmatised', judging from contemporary grammars, a comment on the hypothesis that heavily stigmatised variants are rare in trial transcripts is not possible for the time being. However, one could argue that the frequent occurrence of HAVE TO, originating in conversation and a change from below, shows that at least more informal, if not stigmatised variants, were acceptable in court and in trial proceedings.

The analysis has also shown that the OBC's trial proceedings are worth a look for the investigation of changes in preferred (semi-)modals, as the variants are frequent and can be analysed in quite some detail. However, any analysis needs to take into account that trials provide a particular kind of text (mainly past narrative) and that different trial participants may produce different types of language. The following Chapter 5 takes a look at another Late Modern change: the shift from auxiliary BE to auxiliary HAVE with past participles of mutative intransitive verbs.

5 Auxiliary variation: BE and HAVE with perfects of mutative intransitives

> FRANCIS FREEMAN: […] Martin said, "You **have come** for your money **are you**? I don't much approve of my bargain […]". (OBC, t18200217-20)

The present chapter focuses on variation between BE and HAVE as temporal/aspectual markers with past participles of mutative intransitives (e.g. *you are come* vs. *you have come* in the quotation above), the so-called "*be*/*have* paradigm" (Rydén 1991: 346).[107] While only HAVE is available to signal perfect meaning in contemporary English, both BE and HAVE could fulfil this function in the Late Modern period. In the course of this period, however, a crucial change occurred: initially the minority variant, HAVE became the preferred and finally the only possible alternative. The present chapter explores this change.

Section 5.1 reviews previous research on BE/HAVE variation in the history of English, embedding the current study in a larger context and showing the relevance of the Late Modern changes with regard to observed long-term developments. It also introduces factors that previous work identified as relevant for the choice of auxiliary and outlines contemporary grammars' comments on the issue. Some methodological considerations, such as the criteria for examples to be included in the analysis and the coding conventions, are discussed in 5.2. In Section 5.3 the findings in the OBC and CLMET are discussed and compared in light of previous research and the socio-historical context of the period. Section 5.4 concludes this chapter with a summary.

5.1 Previous research and treatment in LModE grammars

This section provides a brief overview of research on BE/HAVE variation. Section 5.1.1 starts on the most general level with the history of the construction BE/HAVE + participle. Section 5.1.2 moves on to BE/HAVE + participle being used to indicate

[107] In Rydén & Brorström (1987), the term BE/HAVE paradigm is used for the syntactic constructions BE + past participle and HAVE + past participle. The term 'paradigm' is used because these constructions constitute paradigmatic choices, i.e. alternatives.

perfectivity: it charts diachronic change within this paradigm and lists factors that were shown to impact the choice of variant. Section 5.1.3 is dedicated to the treatment of BE/HAVE + past participle in Late Modern grammars.

5.1.1 BE/HAVE + past participle in the history of English

The origins of the BE/HAVE + past participle construction lie in Old English: originally, it contained an adjectival, sometimes inflected past participle and served to denote 'state', as in *hie wæron gecumene* (Rydén & Brorström 1987: 16). As grammatical concord was lost and a fixed word order (with the past participle immediately following the auxiliary) established, BE and HAVE came to be reanalysed as grammaticalised tense-aspect-markers in the course of time (Rydén & Brorström 1987: 16–17). From the Late Middle English period onwards, HAVE gradually became more frequent. However, it was not until the 17th century that HAVE increased more markedly (Rydén & Brorström 1987: 16–17). In Late Modern English, finally, the BE/HAVE paradigm with mutative intransitives changed drastically – from a BE-dominated to a HAVE-dominated paradigm (Rydén & Brorström 1987: 9–10): the general HAVE ratio shifted from ca. 20% around 1700 to ca. 40% around 1800 and finally to ca. 90-95% around 1900 (Rydén & Brorström 1987: 213).[108] Today, only HAVE can be used to indicate perfect meaning.

Originally, HAVE was only associated with transitive verbs (Kytö 1997: 18), but its use soon extended to other verb types. As early as the Old English period, HAVE emerged as an alternative to BE for intransitives. A Middle English example of an intransitive verb in combination with HAVE is shown in (41).

(41) I recomand me to yow, letyng yow weete that I **have spoken** with Herry Colett and entretyd hym in my best wyse for yow, […]
(John Paston II, to John Paston III, 1477, Paston Letters, Vol 1, 504; quoted from Halas 2012: 227)

Among the intransitive verbs, the encroachment of HAVE was slowest among the mutative intransitives, i.e. "verbs denoting some kind of change (positional or otherwise)" like *come*, *change* or *return* (Rydén & Brorström 1987: 9). Although HAVE was slowly being introduced with mutative verbs in late Middle English, the clear

[108] For further references on the subject, see Rydén & Brorström (1987: 16).

majority option was BE (Kytö 1997: 17–18, also see Rydén & Brorström 1987: 16–18). Even in Early Modern English, when HAVE had already prevailed with stative intransitives (like *lie*, *rest*, *stay*), mutative verbs still mainly occurred in combination with BE (Kytö 1994: 184). It was only in the early 1800s that HAVE became more frequent than BE (Kytö 1997: 70). This development, with only small differences in terms of the pace of change, has been observed for British English (see Rydén & Brorström 1987 or Kytö 1997), American English (Kytö 1997 for the long-term development and Anderwald 2014b for the Late Modern period) and Irish English (McCafferty 2014).

While earlier accounts conceptualise this change as a replacement of BE by HAVE that started in late Middle English but only came to completion in Late Modern English (e.g. Rydén & Brorström 1987, Kytö 1997), more recent research proposes that the replacement of BE in fact did not take place until Late Modern English (McFadden & Alexiadou 2010: 417). The Middle English developments are analysed as an expansion of HAVE + past participle at the expense of the simple past and thus as a separate change that is unrelated to later variation between the auxiliaries BE and HAVE (McFadden & Alexiadou 2010: 422).

McFadden & Alexiadou (2010: 415) assume that constructions with BE and HAVE both started out as statives built around resultative participles, but then developed differently. Constructions with BE, as in (42), remained restricted to perfects of result.[109]

(42) I […] wil build againe the Tabernacle of Dauid, which *is fallen* downe. (KJNT,XV,1A.1000; quoted in McFadden & Alexiadou 2010: 401)

HAVE + participle, however, acquired "the full range of interpretations characteristic of the ModE perfect", i.e. developed into a more general perfect, by the end of the Middle English period (McFadden & Alexiadou 2010: 392). The construction starts to appear with readings that are nonresultative, such as the experiential reading in (43).[110]

[109] Perfects of result describe "a state holding at the reference time that is the result of the eventuality described by the verb phrase" (McFadden & Alexiadou 2010: 400).
[110] McFadden & Alexiadou (2010: 400) explain that experiential perfects describe "an eventuality that occurred before the reference time, with no implication that it continues" and provide the following example sentence:
I have been sick twice since January. This implies that the speaker was sick at two separate instances, but – in contrast to a resultative perfect – there is no claim that the speaker is still sick.

(43) For suche as *hath gone* anye tyme abroade, wyll neuer forsake their trade.
'Whoever has gone some time abroad will never forsake their trade.'
(from Thomas Harman's *A caueat or warening* (1567-1568), quoted in McFadden & Alexiadou 2010: 401)

In these contexts, HAVE + participle is not replacing BE + participle, which is incompatible with nonresultative readings in the first place. Rather, HAVE + participle serves as an alternative to the simple past, which would have been used in such contexts in Old English and Middle English (McFadden & Alexiadou 2010: 415, 421). At this point, we cannot yet speak of competition between the two auxiliaries:

> Earlier English does not display a single tense-aspect with an alternation in the auxiliary according to properties of the main predicate and its arguments. Rather, the choice of auxiliary reflects a choice between two distinct temporal-aspectual structures: *have* spells out a Perf head, while *be* is just a copula, accompanying a stative resultative participle.
> (McFadden & Alexiadou 2010: 417)

A similar comment is found in Rydén (1991: 352): although the author considers the change from BE to HAVE as one long process starting in Middle English which is coming to completion in LModE, he remarks that "[f]or a long time, the encoding or realisation of the [BE/HAVE] paradigm was a matter of aspect orientation and feature focussing rather than of true 'perfect' marking", which is close to McFadden and Alexiadou's arguments. However, the latter go one step further and claim that "[t]he actual replacement of *be* by *have* was a separate and later change, which took at most 200 years and was completed around 1900" (McFadden & Alexiadou 2010: 422).

The account involving two separate consecutive changes solves some problems inherent in other explanations. For instance, it provides a convincing narrative on the reasons for the expansion of HAVE. In 'traditional' accounts, the increasing functional load of BE is usually invoked as crucial in the rise of HAVE at the expense of BE (see e.g. Traugott 1972 or Rydén & Brorström 1987): HAVE + past participle, it is argued, presented an unambiguous alternative to BE + past participle, which could be either a passive or a perfect. McFadden & Alexiadou (2006: 258) point out, however, that such ambiguity is only an issue with "verbs that have both transitive and intransitive uses,

which are not distinguished morphologically", an occurrence that was actually "rare in the relevant older stages of the language".

Instead, they link the relative rise in the frequency of HAVE compared to the frequency of BE before 1700 to what they call the 'counterfactual effect': counterfactual clauses, which had been expressed with simple past subjunctive forms in Old English and early Middle English (see e.g. Mitchell 1985: 805), started being expressed with perfects in the first half of the Middle English period "as part of the general expansion of the auxiliary system" (McFadden & Alexiadou 2006: 243). As auxiliary BE was "categorically incompatible with past counterfactual semantics" (McFadden & Alexiadou 2006: 260), this led to a spread of HAVE in counterfactual contexts and thus to the spread of HAVE with the perfect in general (McFadden & Alexiadou 2006: 243).[111] Importantly, this spread of HAVE was not at the expanse of BE, whose frequency remained stable – within its restricted domain – up to around 1700 (McFadden & Alexiadou 2010: 422).

Until this point, BE and HAVE were distributed according to the following rule: "*Be* only forms perfects of result where the result state holds of the subject. […] *Have* appears in all experiential perfects and in perfects of result where the result state holds of something other than the subject" (McFadden & Alexiadou 2010: 399). This is essentially a more rigorous formulation of observations expressed in earlier research, as McFadden & Alexiadou (2010: 399) also acknowledge: Kytö (1997: 31) for instance, describes the distinction between state/result and action as "one of the main distributional factors influencing the choice of the auxiliary", the former being associated with BE, the latter with HAVE.[112]

Conceptualising the pre- and the post-1700 increase of HAVE + participle as effects of two different changes also means that researchers do not have to account for the "unusually long time, i.e. 1,000 years" (Rydén 1991: 352) that the change from BE to HAVE took. Even work that subscribes to the view of a unified 1,000-year change sometimes remarks on the different quality of the developments before the Late Modern period and in the Late Modern period, respectively: Rydén (1991: 343) states

[111] McFadden & Alexiadou (2006: 243) point out that the tendency for modals and counterfactuals to favour HAVE in early English had previously been reported in many earlier studies, including e.g. Traugott (1972) or Rydén & Brorström (1987), but stress that "the tight relationship between the first appearance of such contexts in the perfect and the very first advances of HAVE has not to [their] knowledge been made explicit [before]".
[112] The importance of state and action is also mentioned in Rydén & Brorström (1987: 183).

that "the post-1700 period […] is in many respects the most interesting period and the most crucial one" (1991: 343). In the present work, I follow McFadden & Alexiadou (2006, 2010), i.e. assume that two separate processes took place, of which only the latter is of interest here. Accordingly, I adopt the view that we can only from ca. 1700 onwards speak of auxiliary variation between BE and HAVE and finally a change towards HAVE as the favoured option.

5.1.2 Factors conditioning variation between BE and HAVE

The development of BE and HAVE as perfect auxiliaries has been discussed in various studies, identifying a number of linguistic and extralinguistic factors that correlate with HAVE or BE. A particular focus on the Late Modern period is found in a monograph devoted to the issue by Rydén & Brorström (1987); other corpus studies like Kytö (1997) include Late Modern developments as part of a long-term diachronic exploration of the BE/HAVE alternation in the history of English.

The importance of the counterfactual effect, i.e. that counterfactual semantics require HAVE (McFadden & Alexiadou 2006, 2010), for the expansion of HAVE before the Late Modern period has already been mentioned. In fact, some other contexts only superficially appear 'HAVE-promoting' due to interference from the counterfactual effect. McFadden & Alexiadou (2006: 247) point out that the past perfect, which e.g. Rydén & Brorström (1987: 189) and Kytö (1997: 56) report to favour HAVE, actually shows no independent preference for the auxiliary HAVE. The high rate of HAVE in this context is caused by the counterfactual perfects in their data (which are formally past perfects: *If I had gone to visit her…*), as these categorically take HAVE (McFadden & Alexiadou 2006: 247–248). If they are taken out of the equation, the analysis actually shows a dispreference for HAVE with the past perfect in Middle English and no significant difference between past perfect and present perfect in Early Modern English (McFadden & Alexiadou 2006: 248). However, the preference for HAVE in perfect infinitive constructions, also reported e.g. in Kytö (1997: 56), remains – independently of the counterfactual effect.

Other important linguistic factors found to correlate with HAVE are durative and iterative contexts, certain main verbs and the presence of an object-like complement (Kytö 1997: 70). Some of these are directly connected to the semantics of the BE

perfect: "Iteratives and duratives are about the eventuality expressed by the verb, not its resultant state", which makes them "incompatible with the BE perfect", which is resultative (McFadden & Alexiadou 2006: 253). Among the mutative intransitives, verbs indicating action (typically motion as in, e.g. *arrive, return, enter*) more readily accepted HAVE than those indicating process (typically change of state as in, e.g. *grow,* or *become*), reports Kytö (1997: 36–38).

In general, different main verbs hold on to BE for different lengths of time. While part of the reason must be a verb's potential to express non-resultative meaning, frequency also plays a role. Both Kytö (1997: 45) and Anderwald (2014b: 14) report that COME and GO show the greatest variation the longest.[113] Smith (2007: 260–264) believes that it is no coincidence that these highly-frequent verbs resisted HAVE the longest: Arguing based on a usage-based model of language storage and processing, he assumes that HAVE had a "stronger representation in the mind" of speakers because the use of BE had always been syntactically and semantically more restricted (Smith 2007: 260). Consequently, speakers were more likely to resort to HAVE than to BE when confronted with an infrequent verb as it would not easily be remembered in the BE construction and thus "the more dominant HAVE pattern, which is easily recalled due to its high type frequency, [would] be substituted" (Smith 2007: 262). The only exception would be highly-frequent verbs that are (additionally) strongly associated with the BE pattern. As a result, BE was first replaced by HAVE in constructions with verbs that occurred infrequently (Smith 2007: 261). It has also been observed that BE was retained longer in frequent collocations than elsewhere: *the time is come*, *she is come/returned home* and *he is turned fifty* (Rydén & Brorström 1987: 198). Easily accessible representations in the mind apparently play a role here too.

That the influence of other linguistic factors may have changed over time is at least suggested by findings concerning *-ing* constructions (see (44) and (45)).

(44) My husband **being come** from the pay-table and being a little in liquor, I did not tell him my misfortune that night […]. (OBC, t17520408-19)

(45) she came to look after the children, my father not **having come** back (OBC, t19000212-183)

[113] In contemporary English, *gone* is the only past participle that is still frequently found in combination with BE. It is no longer interpreted as a perfect but rather as an adjective, though, and the alternative HAVE *gone* is available for the perfect (Anderwald 2014: 14).

The finding in Rydén & Brorström (1987: 193) that *-ing* constructions are HAVE-promoting in the 19th century can only partly be confirmed for earlier periods: both Kytö (1994: 188) and McFadden & Alexiadou (2006: 250) report that *-ing* constructions favoured BE in Early Modern English.

Relevant extralinguistic factors are chronology and text type. As outlined above, HAVE became more and more widespread with time. Some text types apparently were more accommodating to this form than others: Rydén & Brorström (1987: 200) report that HAVE spread faster in plays than in letters in their Late Modern data. Kytö (1997: 44–45) notes that the rise of HAVE in journals was ahead of that in other text types (fiction, letter, drama, science, sermon). However, no significant link between HAVE-progression and the level of formality, the degree of orality of a text type or the relationship between a text and the spoken language could be established in her data (Kytö 1997: 49–50).

Whether social factors correlate with perfect auxiliary choice is unclear. In fact, this question could so far not be discussed in detail because existing corpus analyses of the phenomenon only had access to certain groups' language use: Straaijer (2010: 75) remarks that Rydén & Brorström (1987), the most extensive investigation of the BE/HAVE paradigm, only had access to the language of the educated, literate middle classes. Within this group, Rydén & Brorström (1987: 205) suggest that the change was slower in rural communities, citing Jane Austen's preference for BE as an example of rural conservatism. They also remark on "a certain conservatism on the part of women" in general (Rydén & Brorström 1987: 206). However, they are hesitant to assume any social component, e.g. an association with the usage of a particular group or notions of prestige, with any variant. The relative underuse of the incoming variant HAVE is explained as a consequence of a lack of "paradigmatic exposure", i.e. exposure to both variants (Rydén & Brorström 1987: 206). More broadly, the Late Modern change from BE to HAVE is seen as "accelerated by situational (period-inherent) factors like social instability and mobility" from the late 18th century onwards (Rydén & Brorström 1987: 214).

5.1.3 Late Modern grammars on BE/HAVE + past participle

The treatment of auxiliary choice with participles in Late Modern grammars is characterised by two interrelated tendencies: generally rather harsh criticism of BE + participle and considerable terminological and descriptive confusion surrounding this construction.

The 18th-century grammars listed in Sundby et al. (1991: 180–181) contain more critical remarks on BE + participle than on HAVE + participle: out of the 187 grammars surveyed, 14 contain critical remarks on BE, but only four grammars criticise the use of HAVE. Where BE is criticised, HAVE is suggested as a correct alternative in five grammars – the first time in 1766. Similarly, three of the four critical remarks on HAVE come with the recommendation to use BE instead. While this survey of grammars is neither exhaustive nor focussed on the use with mutatives, it nevertheless shows that most 18th-century sources considered HAVE more acceptable than BE.

There was, however, no consensus on one 'correct' alternative. In their examination of the treatment of BE/HAVE with mutatives in 50 18th and 19th century grammars, Rydén & Brorström (1987: 208) find that the choice of auxiliary is often explained as being dependent on the opposition between state and action: BE should be used for statal aspect, and HAVE to indicate perfectivity. Some authors believed that the most appropriate auxiliary depended on context (Straaijer 2010: 67): Priestley (1768: 127–128), for instance, distinguished between contexts of immediacy or 'recentness', in which he deemed BE most appropriate, and contexts of duration or 'pastness', where HAVE is advocated.

For the 19th century, the treatment of BE/HAVE + participle in grammars has been extensively covered based on the Collection of Nineteenth-Century Grammars in Anderwald (2012), Anderwald (2014b) and Anderwald (2016). The phenomenon is quite frequently commented on, with about half of all surveyed grammars mentioning it (Anderwald 2014b: 19). In the early 19th century, variability between BE and HAVE is widely acknowledged (Anderwald 2012: 40). In the following, though, BE falls out of favour: Anderwald (2012: 40) locates the first comments actively preferring HAVE in the 1820s. In the following decades, BE is then explicitly discouraged (Anderwald 2012: 41).

In my own analysis of 16 selected 19th century grammars, a mixed picture emerges: 10 out of 16 grammars accept that BE and HAVE can both be used with the past participle, but that BE is restricted to verbs of motion. Sometimes this has to be inferred from the examples provided, e.g. when Pinnock (1830: 117) states that BE is possible with "certain intransitive verbs" and gives the examples *I am risen* and *I am fallen*. Other authors point to verbs of motion in general (e.g. Allen 1824: 26, Allen & Cornwell 1841: 152, Dawnay 1857: 69, Curtis 1876: 52). A more restrictive remark is found in Higginson (1864: 42), where BE is considered an appropriate auxiliary for the participles of GO and COME, as well as occasionally "other verbs of motion" that are not specified. These mentions of restrictions reflect the diminishing use of BE. Six grammars accept only HAVE + participle for the formation of the perfect. One grammar in the sample, Beard (1854), explicitly discourages the use of BE based on his analysis of all BE + participle constructions as passives: as intransitive verbs cannot form passives, he consequently considers all constructions of BE in combination with participles of intransitives unacceptable.

As Anderwald (2014b: 25) also points out, this misclassification of BE + participle by grammarians is at the root of the comparatively harsh criticism levelled at the construction:[114] at the beginning of the 19th century, present-tense BE + participle was rarely identified as a perfect but usually "described (inadequately, compared to linguistic reality) as the passive of a neuter verb" based on the traditional distinction between active, passive and neuter[115] verb types taken over from medieval Latin grammar writing (Anderwald 2014b: 29). The strong criticism levelled at the BE perfect is "explicitly linked in many cases to the analysis of this form as an improper passive" (Anderwald 2014b: 28). It was only in the 1860s that most British grammar writers adopted the descriptively more appropriate distinction between transitive and intransitive verbs (Anderwald 2014b: 29). But by then, BE was already quite infrequent in usage and associated with "antiquated use" (Anderwald 2012: 41). In light of the fact that HAVE became the majority option for mutative intransitives in the first

[114] It is worth noting that the very arguments that were put forward in favour of HAVE (reason, analogy, regularity in the system) apparently held little weight in the comments on some other changes (Anderwald 2012: 41): for instance, the passival (*the bridge is repairing*) was vehemently defended as superior to the passive progressive (*the bridge is being repaired*), which would have been the more 'regular' alternative, on the grounds that the passival was used by 17th and 18th century authors, i.e. sanctioned authorities (39).

[115] The distinction between these categories was basically semantic, with neuter verbs designating neither activities (as active verbs do) nor the undergoing of an activity (as passive verbs do), but representing states of being (Anderwald 2014b: 21–22).

decades of the 19th century (Rydén & Brorström 1987: 196), research indicates that prescriptions simply reflected usage for this phenomenon. A direct influence of prescription on usage can practically be ruled out.

5.2 Methodological considerations

The present section outlines the key methodological considerations in this chapter, namely how BE/HAVE + mutative intransitives were extracted from the corpora, how relevant examples were identified (largely in line with the procedure outlined in Kytö 1997) and according to which criteria the coding was undertaken.

The present study concentrates exclusively on variation in the use of the auxiliaries BE and HAVE with past participles of one type of verb, namely mutative intransitives (MIs). They represent the only group of verbs still exhibiting widespread variation between BE and HAVE in Late Modern English. With stative verbs, HAVE "already prevails" in the Early Modern period, which constitutes a good reason to focus on mutatives only in variationist studies of later periods (Kytö 1994: 184). For practical reasons, I concentrate on the 10 MIs whose past participles are most frequent in the OBC and that are described as still showing significant variation in Late Modern English in Rydén & Brorström (1987): APPEAR, COME, ENTER, GO, GET, PASS, RETURN, RUN, SET, TURN.

All past participles of these verbs in combination with relevant forms of HAVE and BE (i.e. past and present tense paradigms of the verbs as well as the *-ing* form and the base form) were extracted from the OBC and the CLMET. Variant spellings were taken into account: for the past participles of RETURN, for instance, the search included both *return'd* and *returned*, plus capitalised variants. Where multiple competing participles were in use, this was also accounted for: RUN includes *run* and *ran*, and GET includes *got* and *gotten*.

The search terms by default exclude double perfects (46) and the second elements in coordinated participles (47). This is in line with the methodology described in Kytö (1994) and Kytö (1997). For practical reasons (ease of retrieval), the search also excludes discontinuous BE/HAVE + participle (48).

(46) My Daughter had been out at Service, and **had been come** Home but three Days. (OBC, t17430413-1)

(47) ANNE HARROLD: […] Yes, he has come and **gone** often to my house backwards and forwards (OBC, t17780603-63)

(48) Jane Hatchet. Mrs. Box […] told me Christian Streeter had been gone out to take a walk in the Park, and **was** not **come** home; and she was afraid some ill had come to her. (OBC, t17570420-42)

So-called 'double perfects' (HAVE *been* + past participle) were a third option in addition to BE + past participle or HAVE + past participle at the time. They were first used in the 14th century and disappeared after the 1850s (Rydén & Brorström 1987: 24–26, Denison 1993: 361, 363, Kytö 1997: 30). As they are a marginal type, they are not considered in the analysis. Second elements in coordinated participles are left out because they do not truly have an auxiliary of their own, so to speak: instead, the choice of auxiliary for the coordinated participles may have been influenced by the first element (Kytö 1997: 30).

In principle, reduced forms of the auxiliaries (as in (49)) are included in the search.

(49) Thank God you'**re come** home alive. (OC, t17320114-41)

However, two contracted forms are ambiguous (*'s* and *'d*) and thus cause problems in the analysis: *'s* could stand for *has* or *is* (Kytö 1994: 180–181), and *'d* could theoretically be the reduced version of *had* or *would*. Co-textual information was used to disambiguate between the two options in examples with *'d*. Thus, all instances in which *'d* was a reduced form of *had* were included in the results. However, the form *'s* (50) remains ambiguous:

(50) She**'s gone** to Islington. (OBC, t17250827-33)

As a consequence, instances with *'s* were excluded from the list of results.

To ensure that only contexts in which variation is regularly observed are part of the dataset, all transitive uses of verbs on the list were removed as these show no variability, but are firmly associated with HAVE (Kytö 1997: 18). This covered examples like (51), where *run (back)* is used transitively.

(51) Q. Did the Deceased strike him, or only put him away?
Montgomory. […] he (the Deceased) was at Work, and had a

> little Saw in his Hand, and when he **had run** the Prisoner back, he gave him a little Stroke on the Back, with the Flat of the Saw, - the Stroke would not have kill'd a Fly. (OBC, t17370420-40)

In addition, all non-mutative uses of the verbs in question were removed: *have got*, the combination of HAVE and the participle of GET, can signal possession (*He has got a job*) instead of movement (*He was/had got to London shortly after midnight*). In line with the procedure in Kytö (1997: 29), examples with object-like complements (52) are included in the analysis, although these favour HAVE.

> (52) GEORGE READ: […] he said "You must come in the morning"—I said "No, I **have come** a long way; I can't come in the morning […]". (OBC, t18720108-144)

They are coded for the presence of an object-like complement so that the impact of its presence can be assessed.

It is furthermore important to ensure that any example with BE and HAVE "conveys the notion of perfectivity" (Kytö 1994: 180) in order to guarantee comparability. Consequently, all examples in which BE fulfils other functions, i.e. serves either as a passive auxiliary or as a copula, should be removed from the dataset. This can be tricky because there are ambiguous examples that could either be an active (perfective) use or a passive use of verbs that can be either transitive or intransitive (*he is changed* 'he has become different' vs. *he is changed* 'he has been made different') (Kytö 1997: 28). In the present study, an example was excluded from the analysis where a passive use could be clearly established from the context. In (53), the suspects "were got" to the Compter in the sense that they seem to have been dragged there with some difficulty.

> (53) [The suspects] were afterwards conveyed to Giltspur-street Compter, but they were very unruly both of them, and it was with very great difficulty that they **were got** there. (OBC, t18000219-67)

This is why the example was discarded. Where a passive reading was not justifiable, as in (54), the example was retained.

> (54) I forgot my own safety, and when I **was got** about a yard into the alley; I turned back to see if they [alleged robbers] stop'd these two gentlemen, intending if they had so done, to call the watch […]. (OBC, t17510116-43)

A more difficult issue is the differentiation between BE as copula (*he is changed* 'he is different'), and BE as perfect auxiliary (*he is changed* 'he has become different') (Kytö 1997: 28, Rydén & Brorström 1987: 24). In contemporary English, the only combination of BE + participle that is still frequently found is BE *gone*, and it is indeed analysed as a construction involving a copula and a participial adjective, i.e. with the meaning 'be absent' (Anderwald 2014b: 17). During the Late Modern period, this was not as straightforward. For *he is gone*, for example, there was ambiguity between the older activity reading 'he has gone somewhere' and the adjective reading 'he is (now) absent', especially when there was no further complementation (Anderwald 2014b: 17). In the end, I largely adopted Kytö's (1997) policy at this point: acknowledging the difficulty of judging differences between stative and perfective participles, she considers all instances of dynamic intransitives valid examples "even though the construction in some examples may come closer to a stative than a perfective meaning" (Kytö 1997: 28). It needs to be kept in mind therefore that the present study – as well as other studies operating on similar guidelines – still includes a good deal of ambiguous examples and thus may potentially overestimate the proportion of BE + participle.

All relevant examples of BE/HAVE + participle in the OBC and the CLMET were coded for the linguistic factors mentioned in Table 22.

Factor	**Levels**
AUXILIARY	BE HAVE
MAIN VERB	APPEAR, COME, ENTER, GO, GET, PASS, RETURN, RUN, SET, TURN
COUNTERFACTUAL	Yes No
PRESENCE OF OBJECT-LIKE COMPLEMENT	Yes No
STRUCTURE	present perfect (PresP) past perfect (PastP) perfect infinitive (PerfInf) *-ing* construction (Ing)
PERIOD	1720-1769 1770-1819 1819-1869 1870-1913

Table 22. Coding for analysis of BE/HAVE + participle

In accordance with McFadden & Alexiadou (2006: 244), COUNTERFACTUAL examples are defined as "those clauses where the implication is clearly that the proposition being considered does not (or did not) hold": apart from both the antecedent and consequent clauses of counterfactual conditionals (exemplified in (55)), this includes two further types, namely "clauses which have essentially the function of the consequent of a counterfactual conditional, but have no conditional antecedent" (for instance clauses with *else* as in (56)), and counterfactual wishes (57).

(55) If she [a ship] **had come** down alongside the Strathclyde she **could have saved** many lives. (OBC, t18760403-293)

(56) I am just come to Town and have but 9 d. else I **would have come** up and drunk with you. (OBC, t17360721-6)

(57) Well, I wish he **had come** down himself. (OBC, t18590131-296)

Their important role in the expansion of HAVE has been mentioned in 5.1.1 and 5.1.2. OBJECT-LIKE COMPLEMENTS were defined rather liberally, in line with the procedure in Kytö (1997: 59–60): the category includes complements (to the subject) in the narrow definition of the term 'complement', such as the adjective phrase underlined in (58) or the noun phrase in (59), as well as constructions that would be functionally analysed as adverbials, as in (60).

(58) I hear her hair has turned <u>quite gold</u> from grief. (CLMET3_0_3_262.txt)

(59) he had been concerned in a robbery with people there, stealing lead off a house, and he turned <u>informer</u> (OBC, t17870523-98)

(60) When I was come <u>within 20 Yards of her</u>, I asked her asked her if the Man had not robb'd her? She said, yes. (OBC, t17400416-19)

Linguistic STRUCTURE distinguishes between the present perfect (*they are/have come*), the past perfect (*they were/had come*), *-ing* forms (*being/having come*), and perfect infinitives (such as (61) or (62))

(61) Henry Salter. About eleven or twelve weeks before the deceased died he came to my house; I thought he **might be run** away from his master [...]. (OBC, t17450424-33)

(62) Martha Hopkins. I intended **to have gone** home that Night, but I was seized with a violent cold shaking, soon after I got into the [public] House. (OBC, t17391205-26)

In addition, the OBC data contain automatically extracted information on extralinguistic factors like the gender or social class of the speaker.

5.3 Findings and discussion

This section presents the findings on BE/HAVE variation in the OBC and the CLMET-drama. An overview of Late Modern developments and influential factors is provided in 5.3.1 for the OBC and 5.3.2 across both corpora. A conclusion is offered in 5.3.3.

5.3.1 BE/HAVE variation in the OBC

Table 23 shows the frequencies for the auxiliaries BE and HAVE in the OBC, listed by individual verb.

MAIN VERB	BE	HAVE	total	% of BE
APPEAR	0	52	52	0.0%
COME	369	1,426	1,795	20.6%
ENTER	0	29	29	0.0%
GET	280	940	1,220	23.0%
GO	4,758	1,804	6,562	72.5%
PASS	0	244	244	0.0%
RETURN	11	51	62	17.7%
RUN	66	277	343	19.2%
SET	0	14	14	0.0%
TURN	32	85	117	27.4%
	5,516	**4,922**	**10,438**	**52.8%**

Table 23. Overview of auxiliary choice for MIs under investigation in the OBC

If one looks at all extracted instances combined, BE and HAVE are roughly equally distributed: 53% of the mutatives occur in combination with auxiliary BE, 47% with HAVE. However, it is also evident that individual verbs have very different profiles: while some exclusively select HAVE (APPEAR, ENTER, PASS, SET), one verb, GO, stands out as primarily occurring with BE (72.5%). In addition, the verbs differ greatly in terms of their frequency of occurrence in the OBC: SET is only found 14 times as a past participle, while GO tops the list with 6,562 instances, making up around 63% of all extracted examples. To take the realities of the data into account, the coding for MAIN

VERB was amended to a two-level distinction between GO on the one hand and all other verbs on the other, which marks the most salient division in terms of the distribution of BE and HAVE.

Another important insight from a primary inspection of all results concerns counterfactuals: as it turns out, the 456 counterfactual examples in the OBC are exclusively realised with HAVE, as in

(63) […] the Prosecutor would **have run** after [the defendant], but the Brother-in-Law knock'd him down and beat him, while the Prisoner got away. (OBC, t17370907-40)

This shows that what has been observed for earlier periods is also true for Late Modern English: counterfactual semantics require HAVE (see McFadden & Alexiadou 2006, 2010; also 5.1.2). As this leaves no room for variation, counterfactual examples were excluded from further analysis. The amended figures by main verb can be found in Table 24.

MAIN VERB	BE	HAVE	total	% of BE
APPEAR	0	24	24	0.0%
COME	369	1,243	1,612	22.9%
ENTER	0	29	29	0.0%
GET	280	862	1,142	24.5%
GO	4,758	1,709	6,467	73.6%
PASS	0	234	234	0.0%
RETURN	11	42	53	20.8%
RUN	66	235	301	21.9%
SET	0	13	13	0.0%
TURN	32	75	107	29.9%
	5,516	**4,466**	**9,982**	**55.3%**

Table 24. Overview of auxiliary choice for MIs under investigation in the OBC (excluding counterfactual examples)

The distribution by verb and the overall distribution of HAVE and BE across all 10 mutatives are not substantially affected by the exclusion of counterfactuals.

In accordance with the procedure in 3.5, a logistic regression model including the predictors COMPLEMENT, STRUCTURE, VERB (GO vs. other), SOCIAL CLASS and PERIOD was fit. Details can be found in Table 25: the estimates in the second column indicate the log odds of HAVE + participle.

	estimate *b*	SE	z value	p-value	confidence intervals	
					2.5%	97.5%
Intercept	-3.69230	0.38260	-9.650	<0.001	-4.4388464	-2.9364159
COMPLE-MENT=yes	1.99169	0.09825	20.272	<0.001	1.8009742	2.1862365
STRUCTURE= perfect infinitive	2.09866	0.40210	5.219	<0.001	1.3037673	2.8827746
STRUCTURE= present/past participle	-0.94882	0.36191	-2.622	<0.01	-1.6701529	-0.2482239
MAIN VERB= other	2.60910	0.09435	27.653	<0.001	2.4261582	2.7961233
PERIOD= 1770-1819	1.42463	0.12824	11.109	<0.001	1.1752951	1.6781770
PERIOD= 1820-1869	2.77474	0.13444	20.640	<0.001	2.5143077	3.0414347
PERIOD= 1870-1913	4.69672	0.16195	29.001	<0.001	4.3834574	5.0184201
CLASS= lower	-0.34121	0.08035	-4.246	<0.001	-0.4988851	-0.1838298
Concordance Index *C*		0.93				

Table 25. Output of logistic regression including predictors COMPLEMENT, STRUCTURE, MAIN VERB, PERIOD, CLASS; based on OBC

The factor GENDER, which was also coded for, did not contribute significantly to the explanatory power of the model and was therefore dropped.

The linguistic factors influencing the distribution of BE and HAVE are the presence of a COMPLEMENT, the verbal STRUCTURE, and the MAIN VERB. As expected based on earlier findings (see e.g. Kytö 1997), the presence of a complement positively impacts the chances of HAVE occurring.

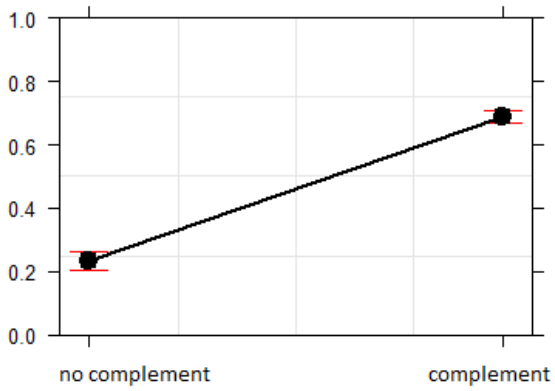
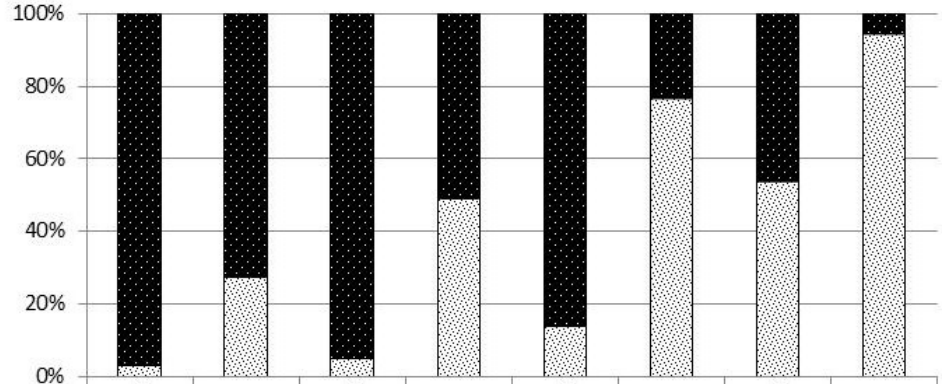


Figure 20. Effect of factor COMPLEMENT on likelihood of HAVE + participle (data: OBC)

In Figure 20, the probability of HAVE occurring with a complement is at ca. 68.8%, as opposed to 23.2% without a complement. According to Kytö (1994: 182), having a

complement in the utterance emphasises the element of 'action', which promotes the use of HAVE. The observed frequencies by period in Figure 21 show that the proportion of HAVE is consistently higher in all periods when a complement is present.



	no comple-ment	comple-ment	no comple-ment	comple-ment	no comple-ment	comple-ment	no comple-ment	comple-ment
	1720-1769		1770-1819		1820-1869		1870-1913	
BE	912	1,239	930	901	830	452	179	73
HAVE	29	468	50	875	137	1,465	208	1,234

Figure 21. Variation between BE and HAVE as perfect auxiliary in contexts with and without complements, by period, OBC (N = 9,982)

In contexts without complements, HAVE only becomes the majority choice in 1870-1913 (53.7%). In environments with complements, a proportion just shy of the 50%-mark (to be precise, 49.3%) is already found 100 years earlier (1770-1819).

Figure 22 illustrates the importance of the effect of the linguistic structure.

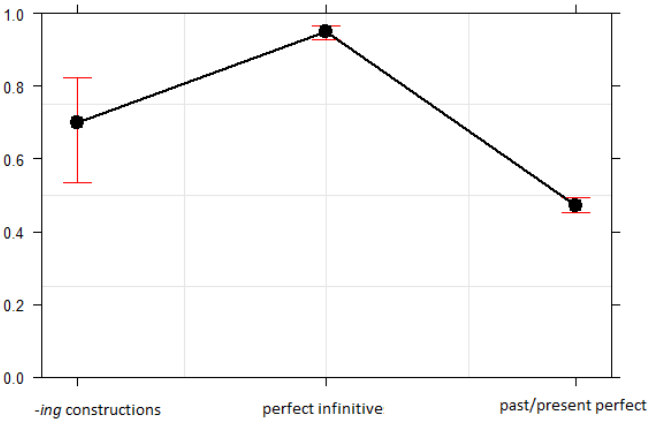
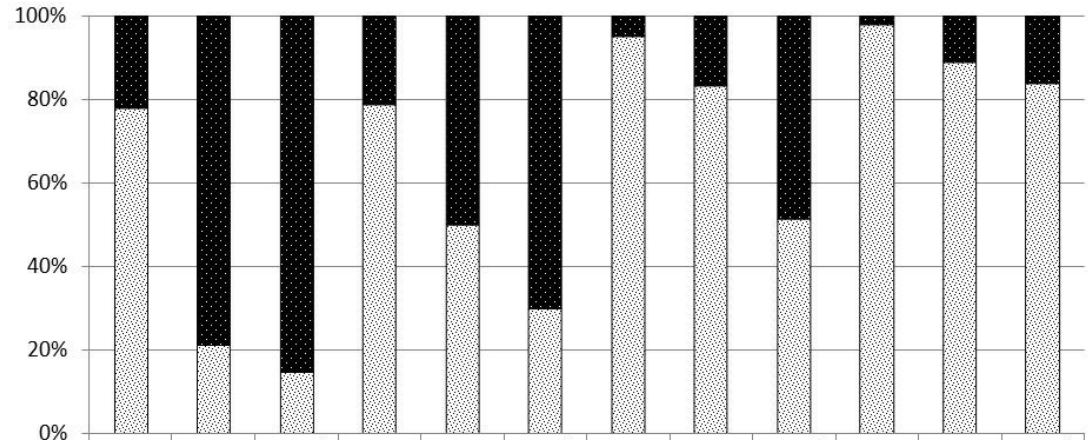


Figure 22. Effect of factor STRUCTURE on likelihood of HAVE + participle (data: OBC)

In this figure, as well as in the summary of the model in Table 25, only three levels are distinguished: past/present perfect, *-ing* constructions and perfect infinitives. The levels past perfect and present perfect were conflated because the analysis did not reveal any significant difference between these contexts. This is not entirely unexpected: McFadden & Alexiadou (2006: 248) had reported the same for the Early Modern period.

The probability of HAVE occurring is highest for perfect infinitives (95.0%), followed by *-ing* constructions (69.9%). The playing field is more even in past or present perfect constructions, where the distribution of BE and HAVE is close to 50-50. In fact, a very slight preference for BE is found in these contexts (chance of HAVE occurring: 47.3%). That the perfect infinitive is an environment strongly favouring HAVE, remarked e.g. in Kytö (1997: 56), is also confirmed by the OBC results. The highest probability of HAVE occurring is found in these contexts, and it clearly is a contender for a "near-blocking" context of BE, as Rydén & Brorström (1987: 193) suggested. Less restrictive, but still more likely to show HAVE than BE, are contexts with *-ing* forms.

To find out whether there is any diachronic development, the observed frequencies by period are shown in Figure 23.



	1720-1769			1770-1819			1820-1869			1870-1913		
	perfect inf.	-ing constr.	past/ present perfect	perfect inf.	-ing constr.	past/ present perfect	perfect inf.	-ing constr.	past/ present perfect	perfect inf.	-ing constr.	past/ present perfect
BE	35	37	2,079	41	12	1,778	13	4	1,265	3	2	247
HAVE	124	10	363	151	12	762	248	20	1,334	146	16	1,280

Figure 23. BE and HAVE participle, by structure and period, in the OBC

In general, the hierarchy of HAVE-promoting factors holds for all periods. When the percentages of HAVE among the three categories are compared (see Figure 24), a case can be made for *-ing* constructions growing into a HAVE-promoting environment. In Early Modern English, they are still reported to favour BE (Kytö 1994, McFadden & Alexiadou 2006), but Rydén & Brorström (1987: 193) claim that *-ing* forms are HAVE-promoting in Late Modern English.

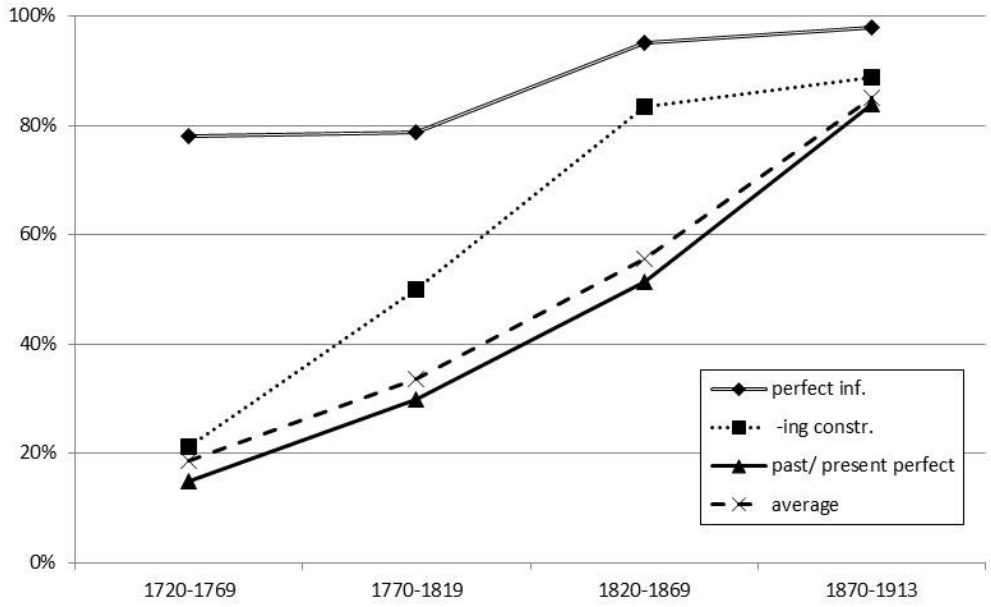


Figure 24. Observed percentage of HAVE by construction and average percentage of HAVE across all constructions, by period, in the OBC

In Figure 24, the strongest increase in the proportion of HAVE is recorded for *-ing* constructions: the slope is much steeper from period 1 to period 3 than in other environments. In fact, the proportion of HAVE rises from 21.3% to 83.3% in that time frame. Nevertheless, the figures indicate that it is not warranted to speak of *-ing* constructions as a "near-blocking context" for BE throughout the entire Late Modern period (Rydén & Brorström 1987: 184–195). Rather, this restriction seems to have developed with time.

The effect of the MAIN VERB is quite clear and consistent in the OBC: The regression model predicts a clear preference for BE with GO and a clear preference for HAVE with the other verbs. Figure 25 illustrates this: HAVE is much less likely with GO (chance of 30.4%) than with other verbs (85.6%).

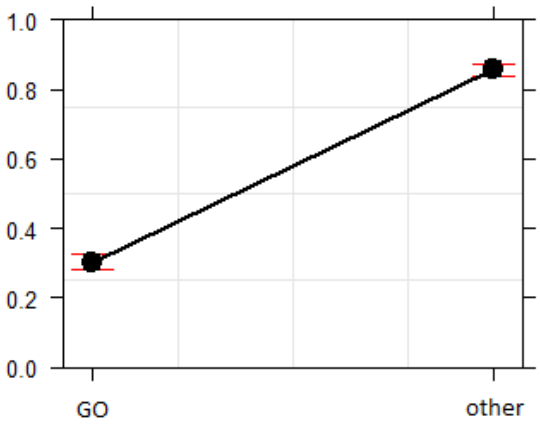
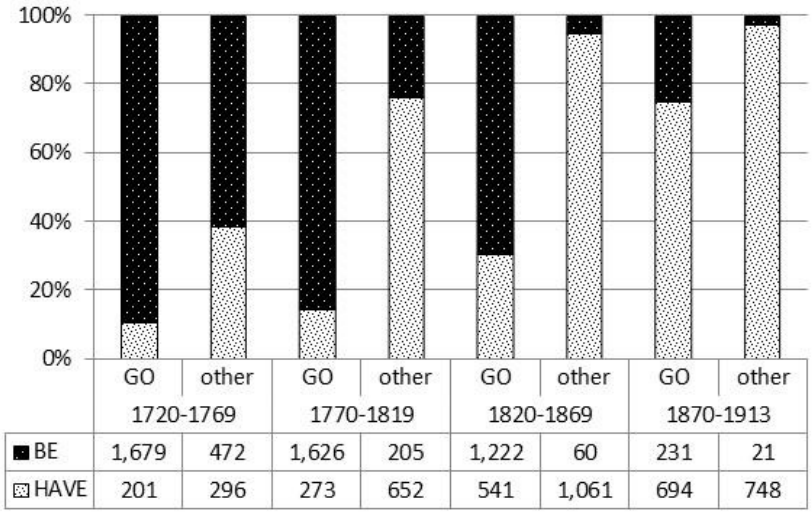


Figure 25. Effect of factor VERB on likelihood of HAVE + participle (data: OBC)

This effect, too, is consistent across time, as the observed frequencies in the OBC show (Figure 26).



	GO	other	GO	other	GO	other	GO	other
	1720-1769		1770-1819		1820-1869		1870-1913	
BE	1,679	472	1,626	205	1,222	60	231	21
HAVE	201	296	273	652	541	1,061	694	748

Figure 26. Variation between BE and HAVE, by verb, by period, in the OBC (N = 9,982)

It is only in the last period (1870-1913) that HAVE is found as the majority option with the verb GO, which remains a stronghold of BE for a long time. This makes good sense, too: as suggested in 5.2, constructions with GO are the only ones still remaining in present-day use, due to the reanalysis of *gone* as an adjective. Since distinguishing the older activity reading from the adjective reading is difficult, the numbers in Figure 26 include both. Additionally, *gone* is also the most frequent participle in the present study. Its continuing association with BE throughout the Late Modern period supports

the theory that high-frequency elements like BE + *gone* are stored as units in the mind and therefore more resistant to change (5.1.2). The individual developments of the most frequent mutative verbs in the study are shown in Figure 27.

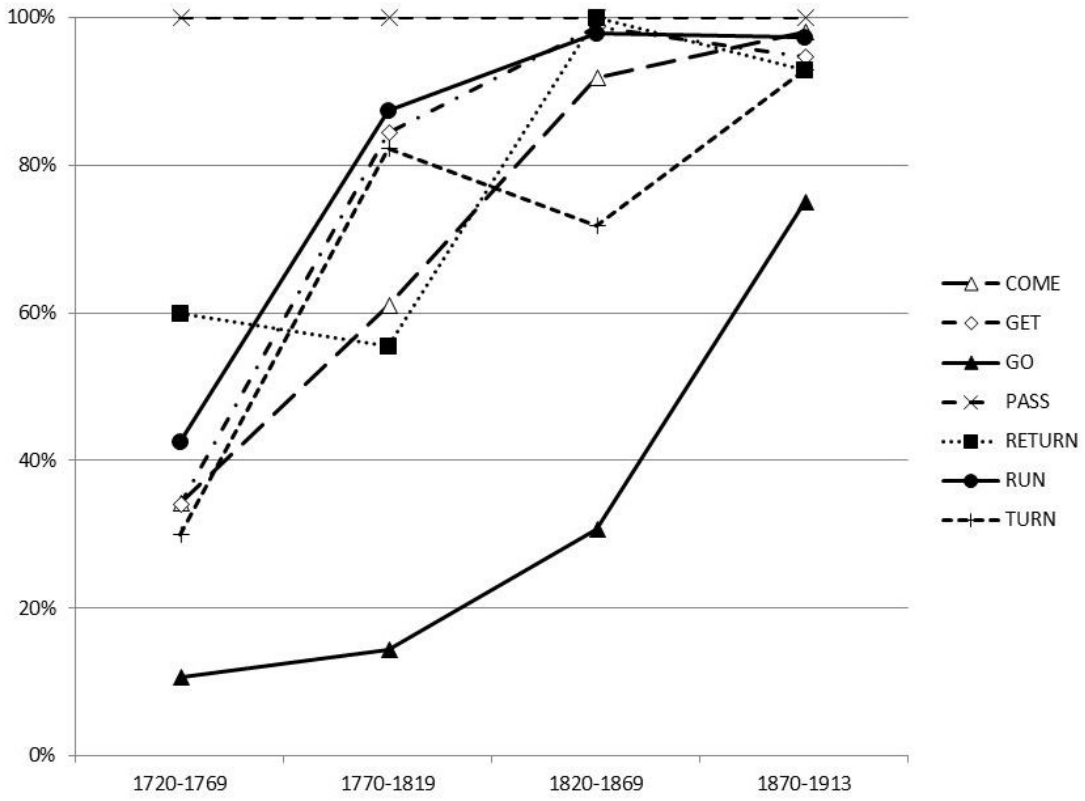


Figure 27. Observed proportions of HAVE, by verb and period, OBC (only for verbs that occur at least five times/period), N = 9,916

It should be borne in mind that these percentages are based on widely differing absolute frequencies (for details, see Appendix: Table A-1) and need to be considered with caution. Overall, though, all verbs except GO cluster in one area. PASS is an interesting case: it always appears in combination with HAVE in the OBC. A clear dominance of HAVE for this verb is also reported in earlier work (e.g. Rydén & Brorström 1987: 133–140). That not a single example with BE is found in the OBC may be due to the semantics of the verb in said examples. Most of the tokens in the OBC (218 out of 242) represent PASS in the sense of either 'happen/take place' and

'move' and thus meanings which were dominated by HAVE throughout the Late Modern period (Rydén & Brorström 1987: 134, 137).[116]

Among the external factors, time and social class play a significant role for this variable. The effect of time, already hinted at in other representations above, is very clear in Figure 28, which contains a graphic depiction of the effect of the factor PERIOD in the regression model: the probability of HAVE rises from 10.3% in the first period (1720-1769) via 32.2% and 64.7% in subsequent periods to 92.6% in 1870-1913. HAVE is gradually replacing BE. The proportions of BE and HAVE in the first period are practically reversed in the final period.

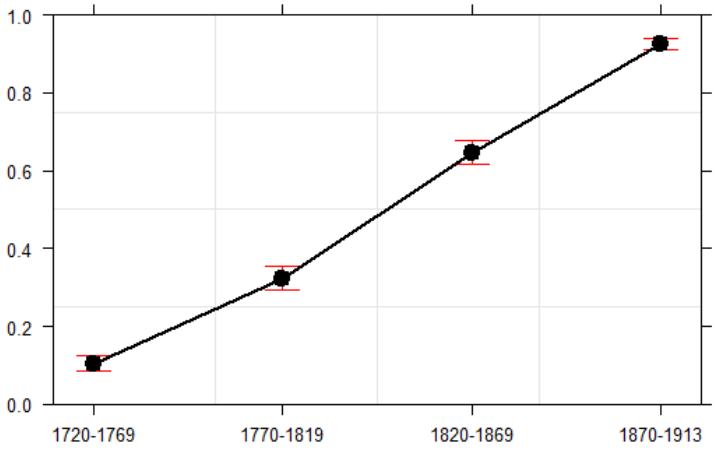


Figure 28. Effect of factor PERIOD on likelihood of HAVE + participle

The decade-by-decade breakdown in Figure 29 shows that the observations for the OBC closely match those in Rydén & Brorström (1987: 196), where the period in which HAVE reaches "paradigmatic majority (+ 50%)" is dated to "the first few decades of the 19th century".[117]

[116] According to Rydén & Brorström (1987: 134, 137), BE was most likely to occur when PASS was used in the sense 'be over'. No unambiguous example in that sense is found in the OBC.
[117] Paradigmatic majority of a variant is reached when the relative frequency of the expansive variant exceeds 50% (Rydén & Brorström 1987: 13).

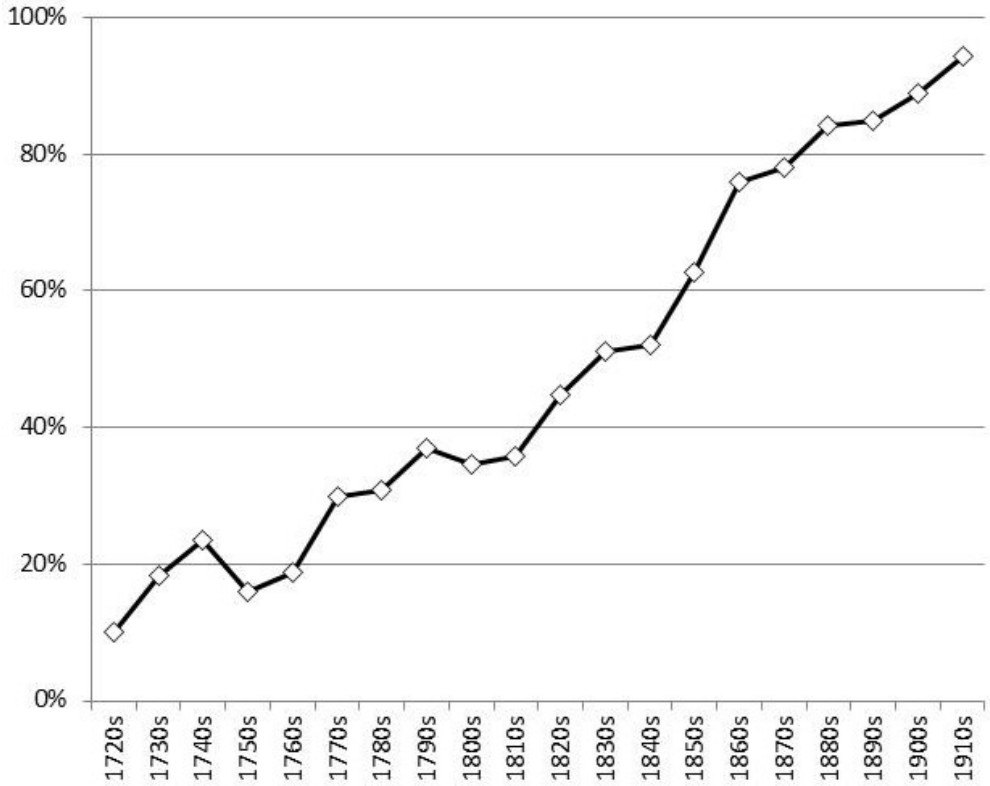
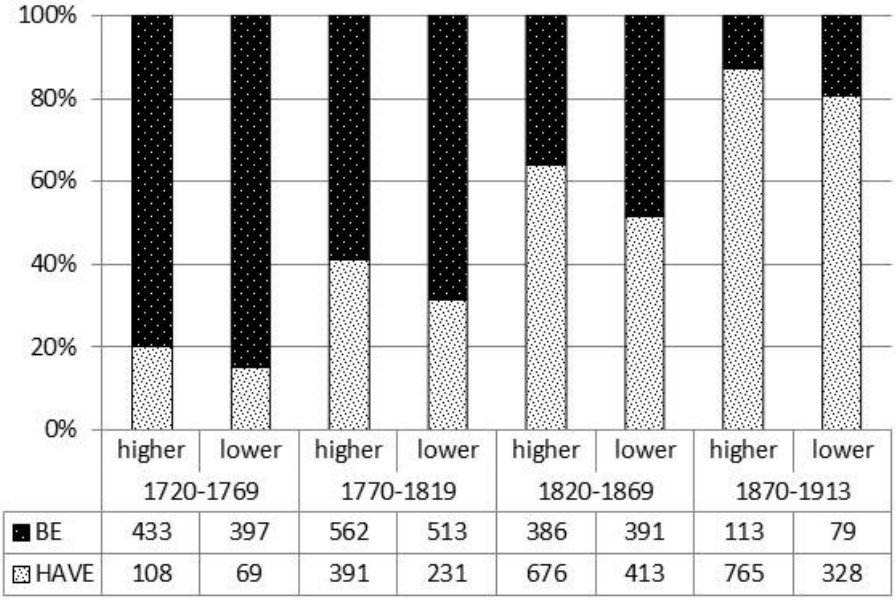


Figure 29. Observed proportions of HAVE, by decade, OBC (N = 9,982)[118]

This paradigmatic majority is reached in the 1830s in the OBC (51.0% HAVE).[119]

Analysis of the OBC data finally shows that SOCIAL CLASS was also a significant predictor for the choice of auxiliary. More concretely, the variant HAVE is consistently found more often in the higher social classes than in the lower social classes. The diachronic development of the distribution of the auxiliaries by speakers' social class is shown in Figure 30.

[118] For absolute frequencies, see Appendix: Table A-2.
[119] It needs to be borne in mind that the analysis in Rydén & Brorström (1987) is based on more verbs than the ten selected verbs in this study, and uses private letters and comedies as source materials. 1:1 comparisons therefore cannot be made.



	higher	lower	higher	lower	higher	lower	higher	lower
	1720-1769		1770-1819		1820-1869		1870-1913	
BE	433	397	562	513	386	391	113	79
HAVE	108	69	391	231	676	413	765	328

Figure 30. BE and HAVE by social class in the OBC (N = 5,855)

The regression model (see Table 25) indicates a small but significant effect of class (Figure 31). The higher social classes marginally favour HAVE (57.4%). Among the lower social classes, the probabilities of BE and HAVE are almost equal (to be exact, the probability of HAVE occurring is 49%, that of BE therefore 51%).

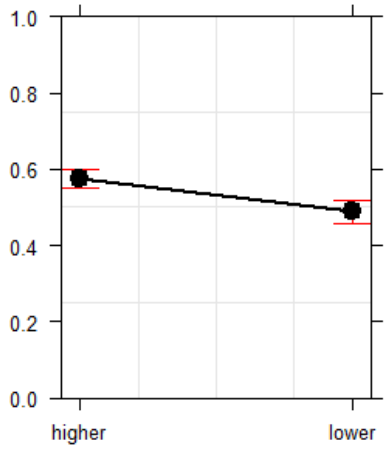


Figure 31. Effect of factor SOCIAL CLASS on likelihood of HAVE + participle (data: OBC)

The higher classes are also ahead of the lower classes in the change from BE to HAVE in each period under investigation. This effect is present throughout the Late Modern period in the OBC. Therefore, it cannot be considered a consequence of normative

grammar-writing, which those higher on the social ladder might have been more exposed to or concerned about. After all, the active preference for HAVE in grammars only starts in the 1820s. It is thus not clear why this class effect holds.[120] This requires further investigation beyond the scope of this study.

5.3.2 BE/HAVE variation in comparison to the CLMET

In order to contextualise the results from the OBC and identify whether different speech-related genres behave differently when it comes to BE/HAVE variation, data from the CLMET-drama is extracted and analysed.

The CLMET-drama contains 619 relevant tokens, 336 for HAVE and 283 for BE. Just as in the OBC, the restriction of BE to non-counterfactual sentences applies, which is why the 36 tokens found in counterfactual clauses were removed from the analysis. Additionally, I decided to exclude all *-ing* clauses from the comparative analysis of CLMET and OBC, as only 3 relevant tokens could be retrieved from the CLMET-drama. The *-ing* category was also the smallest among the STRUCTURE types in the OBC (113 tokens overall for *-ing*), which led to large confidence intervals and less reliable predictions, but the situation in the CLMET is obviously even more problematic. Leaving them in would negatively impact the quality of the regression model.[121] None of the other predictors (COMPLEMENT, PERIOD) caused similar problems. Finally, then, we are left with 580 tokens from the drama corpus, whose distribution is shown in Table 26.

	BE	HAVE
1710-1780	116	41
1780-1850	106	50
1850-1920	59	208
sum	**281**	**299**

Table 26. BE and HAVE, by period, in the CLMET-drama

[120] The effect also holds across different courtroom roles, so this can be excluded as a potential underlying factor. As the factor role cannot be effectively dealt with within the current set-up and methodological framework (see 3.5), it is not further discussed here.
The distribution by role is as follows: defendant: 210x BE, 221x HAVE; judge: 62x BE, 94x HAVE; lawyer: 121x BE; 159x HAVE; victim: 1,641x BE, 782x HAVE; witness: 2,220x BE; 2,798x HAVE (N = 8,308).

[121] In fact, I initially did run a model including the category *-ing*, but this model did not meet the standards of the quality control procedure advocated in Levshina (2015) and outlined in 3.5.

Combined with the reduced set of OBC tokens (without *-ing* constructions: 9,869), the analysis in this chapter is based on 10,449 tokens.

A regression model including the predictors CORPUS, STRUCTURE, MAIN VERB, PERIOD and COMPLEMENT was run. It is summarised in Table 27.

	estimate *b*	**SE**	**z value**	**p-value**	**confidence intervals**	
					2.5%	**97.5%**
Intercept	-5.45704	0.16808	-32.466	<0.001	-5.7893075	-5.1302398
COMPLE-MENT=yes	1.97039	0.07425	26.536	<0.001	1.8259906	2.1171541
STRUCTURE= perfect infinitive	2.94702	0.13110	22.480	<0.001	2.6936612	3.2078648
STRUCTURE= present perfect	0.46754	0.08270	5.654	<0.001	0.3056617	0.6299042
MAIN VERB= other	2.28704	0.06585	34.733	<0.001	2.1588366	2.4170060
PERIOD= 1780-1850	1.48349	0.07035	21.087	<0.001	1.3463751	1.6221987
PERIOD= 1850-1920	3.76110	0.09083	41.409	<0.001	3.5847905	3.9409236
CORPUS= OBC	1.10411	0.13356	8.267	<0.001	0.8428153	1.3665321
Concordance Index *C*		0.91				

Table 27. Output of logistic regression including predictors COMPLEMENT, STRUCTURE, MAIN VERB, PERIOD and CORPUS; based on OBC and CLMET-drama

All predictors emerged as significant to the distribution of BE and HAVE. As the bulk of the data for this model is OBC data, it is not surprising that the same predictors that had already been singled out in 5.3.1 are once again flagged as significant.

What is important to note is the clear difference between corpora that represent different genres. If we consider the effect of the variable CORPUS as predicted by the model (see Figure 32), it emerges that BE is the more likely alternative in both corpora, but the chances of HAVE occurring nevertheless differ: the fitted probability of HAVE is 20.6% for the CLMET and thus much lower than the 43.8% in the OBC.

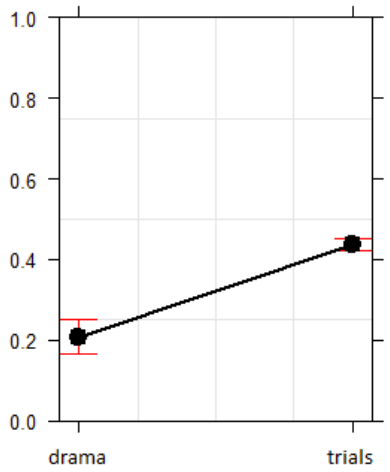
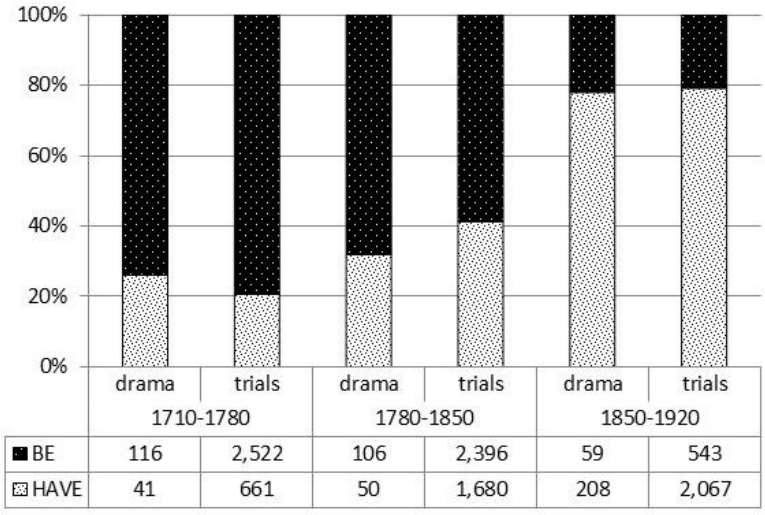


Figure 32. Effect of factor CORPUS on likelihood of HAVE + participle (data from OBC and CLMET-drama)

The development of the observed frequencies can add more depth to this picture (Figure 33):



	drama	trials	drama	trials	drama	trials
	1710-1780		1780-1850		1850-1920	
BE	116	2,522	106	2,396	59	543
HAVE	41	661	50	1,680	208	2,067

Figure 33. BE/HAVE by corpus, by period (N = 10,449)

In both corpora, the overall progression is similar, but the OBC exhibits a slight but significant advantage over the CLMET in terms of HAVE use in the second period.[122] It seems that the fictitious dialogue in the plays is more conservative here than speech in the courtroom. The progression through time in general is also a significant factor, as a depiction of the effect of the factor PERIOD in Figure 34 indicates.

[122] Results of χ^2 test for 1780-1850: $\chi^2 = 4.94$, df = 1, $p<0.05$.

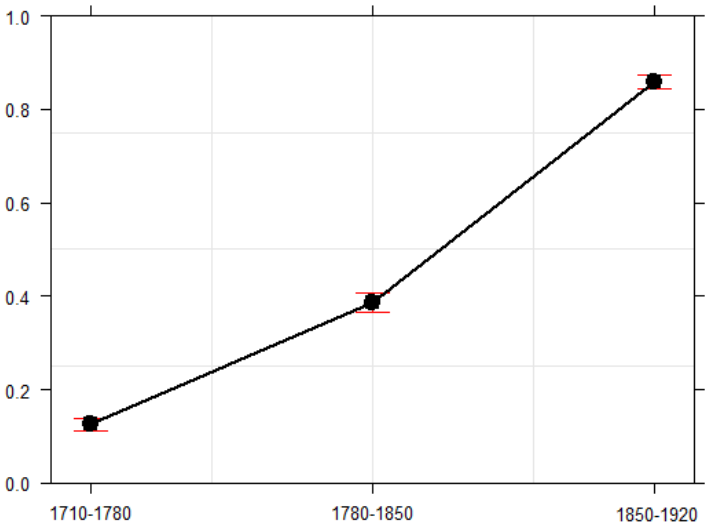


Figure 34. Effect of factor PERIOD on likelihood of HAVE + participle (data: OBC and CLMET-drama)

The probability of HAVE grows significantly between periods 1 and 2 (from 12.5% to 38.7%), and attains a clear majority in the final period (86%).

COMPLEMENT, STRUCTURE and VERB also significantly influence the choice of auxiliary. HAVE is slightly more likely than BE when a complement is present in the context (probability of HAVE: 58.4%; see Figure 35). Without a complement, the probability of HAVE sinks to 16.4% - BE is clearly preferred.

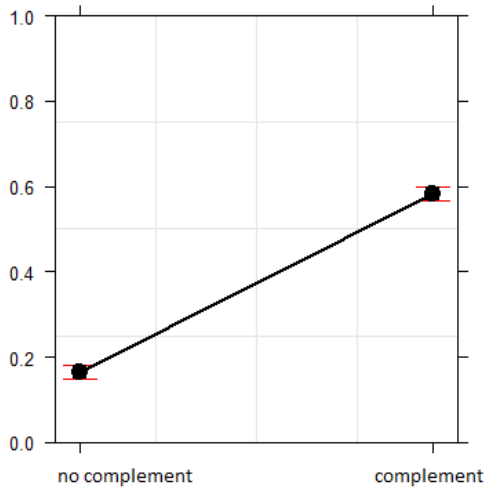


Figure 35. Effect of factor COMPLEMENT on likelihood of HAVE + participle (data: OBC and CLMET-drama)

This confirms the importance of the presence of a complement, which was already shown for the trial texts in 5.3.1. When looking only at the dramatic texts (see Figure 36), the observed frequencies show that the presence of a complement is also associated with a greater use of HAVE in this genre.

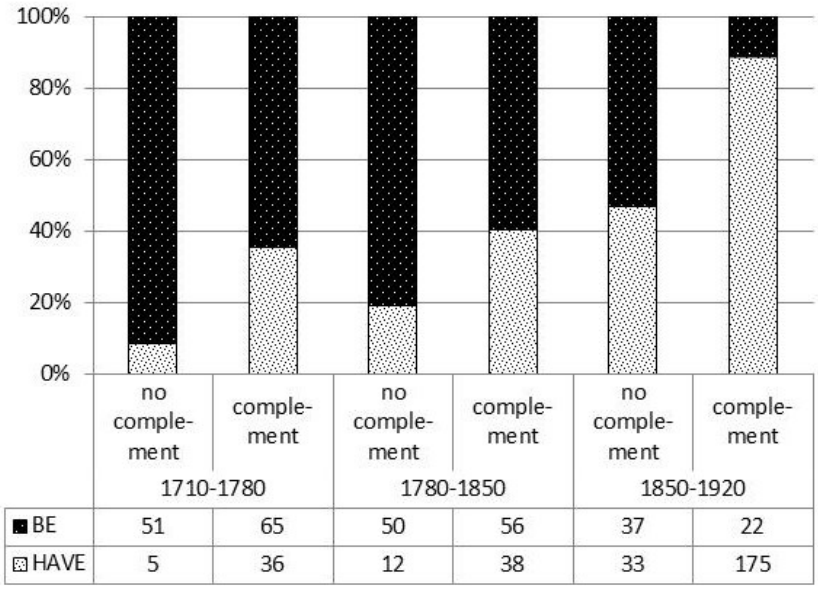


Figure 36. BE/HAVE by complement, by period, in the CLMET-drama (N = 580)

The construction in which the auxiliary is embedded (variable STRUCTURE) also influences the choice between BE and HAVE, as the effect plot for this variable in Figure 37 demonstrates.

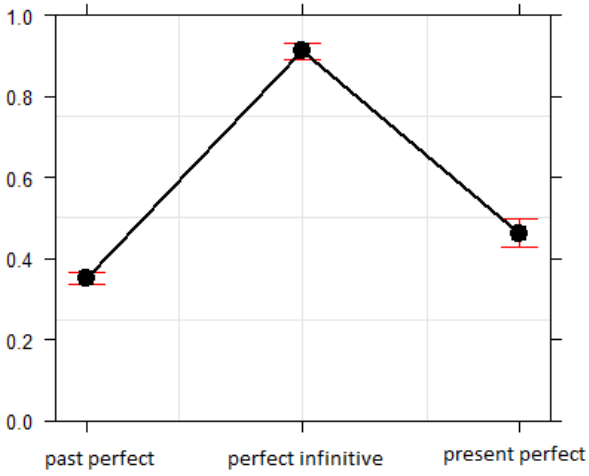


Figure 37. Effect of factor STRUCTURE on likelihood of HAVE + participle (data: OBC and CLMET-drama)

The predicted probability of HAVE is highest with the perfect infinitive (91.1%), while present perfect and past perfect (46.3% and 35.1%) slightly favour BE. It is unfortunate that *-ing* forms do not occur in sufficient quantities to assess whether the CLMET shows the same trend as the OBC, where *-ing* constructions become a HAVE-promoting environment.

In addition to complements and verbal constructions, the verb as such also plays a role in auxiliary choice (see Figure 38): once again, the probability of HAVE is low with GO (24.1%), but high with other verbs (78.7%).

Figure 38. Effect of factor VERB on likelihood of HAVE + participle (data: OBC and CLMET-drama)

That this is also the case independently of the OBC is shown in Figure 39, which displays only the CLMET data by period and verb.

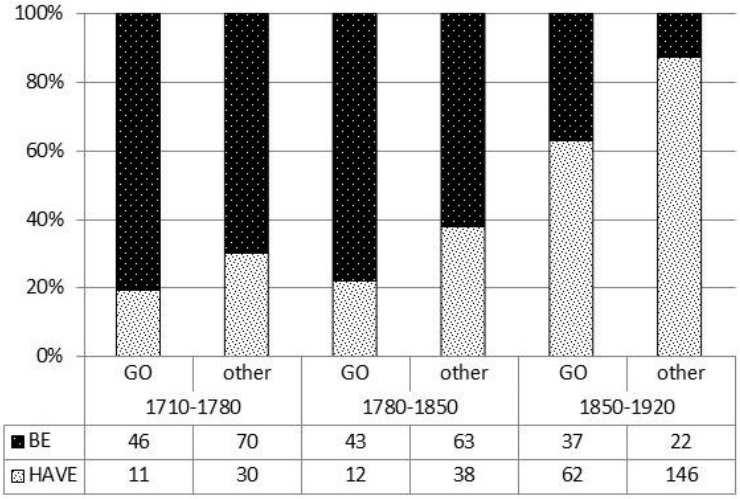


Figure 39. BE/HAVE by verb, in the CLMET-drama (N = 580)

The verb GO remains resistant to the introduction of HAVE: in each period, the percentage of HAVE is noticeably smaller with the verb GO than with other verbs, and the difference between these two groups is growing (1710-1780: 19.3% vs. 30.0%; 1780-1850: 21.8% vs. 37.6%; 1850-1920: 62.2% vs. 86.9%). As already argued for the OBC in 5.3.1, I assume that the high incidence of BE + *gone* instead of HAVE + *gone* is again caused by a) BE + *gone* being stored as a fixed expression and b) verbal and adjectival readings of *gone* being included in the above figures due to the difficulty of distinguishing the two.

5.3.3 Conclusions

This section has investigated BE/HAVE variation in 18th and 19th century English with mutative intransitives. Various factors in the linguistic context have been identified as influential in both the OBC and the CLMET, and the factor SOCIAL CLASS has been shown to impact the distribution in the OBC, with a slightly higher chance of HAVE occurring with higher class speakers. In terms of the temporal development, both corpora show progressions expected based on earlier research, with the proportions of BE and HAVE practically reversing from the beginning to the end of the Late Modern period.

Perhaps the trickiest finding is the social class association of BE and HAVE recorded for the OBC: the variant HAVE is found more often in the higher social classes than in the lower social classes. This is consistent across all periods under investigation, with the higher social classes ahead of the change, and also a general effect (probability of HAVE among the higher social classes: 57.4%; lower classes: 49%). Unfortunately, it is not possible to compare these results to other studies, as there simply are none. Rydén & Brorström (1987) provide the only other historical study of BE/HAVE variation in Late Modern England which addresses social factors to some extent. However, it features only data by the educated middle classes. Based on the present study, the influence of normative grammar can be excluded as a potential explanation; further investigation of this issue will have to follow, though.

More easily interpretable are the findings on linguistic factors. Particularly the verb is important in both corpora, with GO being a stronghold of BE. The presence of a complement improves the likelihood of HAVE drastically in both corpora. The role of

the constructions in which the auxiliaries feature also plays a role. The OBC data allow formulating the following hierarchy of HAVE-promoting environments: perfect infinitive > *-ing* constructions > present & past perfect (levels merged as they do not differ significantly from each other). Combining the OBC and CLMET data, a very similar picture emerges: perfect infinitive > present perfect > past perfect. As *-ing* constructions were practically unavailable in the CLMET and thus had to be excluded from the second analysis, the only difference between these two hierarchies is the more differentiated result for present and past perfect in the combined OBC-CLMET model.

At any rate, these results on BE and HAVE in different constructions support the assertion in earlier research that the past perfect does not independently promote HAVE usage (e.g. found in McFadden & Alexiadou 2006: 247–248): in fact, this construction comes in last place in both hierarchies. Instead, only formal past perfects which are counterfactuals are associated with HAVE. In this study, counterfactuals are a 'knock-out context' for BE.[123] As for the *-ing* constructions, it was unfortunate that a lack of data in the CLMET precluded further investigation into their impact on auxiliary choice. Nevertheless, it could be established that they do not represent a "near-blocking context" for BE throughout the entire Late Modern period, as claimed in Rydén & Brorström (1987: 195). The present results rather suggest that this tendency towards blocking only developed towards the end of the Late Modern period.

5.4 Summary

The present chapter discussed the use of the auxiliaries BE and HAVE with past participles of mutative intransitives in Late Modern English, particularly with regard to HAVE ousting BE as the preferred auxiliary for this group of verbs. After a review of previous research (5.1) and a discussion of key methodological considerations (5.2), Section 5.3 presented the results of a corpus analysis in the OBC and the CLMET. Having reported and considered the corpus results, it is now time to have another look at the hypotheses generated prior to the study (see 3.6.)

Initially, I assumed that a change in progress from the BE perfect towards the HAVE perfect would be visible in the corpora, and that the change would progress at

[123] The term 'knock-out context' is taken over from Aarts, Close & Wallis (2013: 14).

different speeds in the two corpora. These hypotheses can be confirmed. It should be noted, though, that the differences between the two corpora are quite small for this feature: the OBC only exhibits a slightly faster progression from BE to HAVE than the CLMET-drama.

One major methodological challenge in the present chapter was the choice of mutatives to include in the analysis. Obviously, this step has a great impact on all findings, as developments differ for individual verbs. For practical reasons, 10 highly-frequent MIs form the basis of the present analysis. It emerged that differences hold especially between GO, which is strongly associated with BE, and the other verbs. A related issue here is the impossibility to clearly differentiate between truly verbal and (rather) adjectival uses of GO. The present chapter also highlights the importance of rethinking methodology in the course of the analysis. For instance, counterfactuals were initially extracted, but then discarded from the analysis as they never appeared with BE. Similarly, when analysing verbal constructions, the group of *-ing* constructions was excluded because there were not enough tokens for reliable analysis in the CLMET.

After Chapter 4, which discussed the changes from obligative marker MUST to HAVE TO and from past epistemic MUST to MUST *have*, and the present chapter, which dealt with the change from BE to HAVE as the perfect auxiliary for mutative intransitive verbs, the following Chapters 6 and 7 will highlight features that are not involved in change, but assumed to exhibit more or less stable variation in the Late Modern period and in present-day English. We begin with an investigation of historic present and past forms in discourse introducers (Chapter 6 on *I says / I said*) before we move on to a case of subject-verb agreement (Chapter 7 on *you was / you were*).

6 Alternation between historic present and simple past in narrative: Tense of the discourse-presenting verb SAY

> **I said**, you dog, what do you run away for, **says I**, you must have done something, or you would not have run away [...]. (OBC, t17830430-61)

This chapter is concerned with the use of SAY as a discourse-introducing verb in spoken past narrative, in particular whether it appears as past tense *said* or as historic present *says.* A detailed variationist study is conducted on the most frequent combination: *says/said* plus the first person singular pronoun *I*. Compared to the other features discussed in the present work, tense shifting in discourse introducers[124] with SAY is probably the least well-described. A plausible reason is the fact that the historic present is strongly associated with spoken language (Carter & McCarthy 2006: 822) – and thus a kind of language that is difficult to access for historical linguists. It may be encountered in fictional dialogues, but is not frequent in other written texts. The Old Bailey Corpus promises to be a fruitful resource for this matter.

In the following, previous work on tense shifting and contemporary grammars' take on the variation between *says* and *said* is reviewed (6.1). After discussing some methodological issues particular to this case (6.2), Sections 6.3 and 6.4 present findings from the OBC and the CLMET, discussing both linguistic and social factors. As the results from the OBC point towards possible scribal interference, this topic is also addressed. It seems that the OBC only in parts accurately represents speakers' tense choices in the courtroom, and that the changing realities of the courtroom and people's stylistic consciousness need to be taken into account for this feature. The chapter closes with a summary of salient findings and their implications in 6.5.

6.1 Previous research and treatment in LModE grammars

Variation between *I says* and *I said* is located within the broader context of historic present / simple past alternation in narratives about past events. As a consequence,

[124] The term 'discourse introducer' for clauses like *he said*, *she goes*, and *he was like*, which are used to keep track of who is speaking, is taken from Johnstone (1987: 34).

tense-shifting in conversation is reviewed in 6.1.1, before the discussion moves on to tense variation in discourse introducers in 6.1.2. The research review will be a little shorter than for other chapters: no diachronic study on *says/said* variation is known to me, and most general studies on tense shifting, whether in discourse introducers or not, are focussed on the functions of tense switching rather than the social dimension of this conversational feature. Rühlemann (2007) is a notable exception, including discussion of regional, gender and class preferences.

6.1.1 Tense shifting in narrative

Grammatical descriptions of present-day English draw attention to the fact that speakers engaged in conversation routinely shift tenses, typically between present and past tense (see e.g. Biber et al. 1999: 1120, Carter & McCarthy 2006: 360, Huddleston & Pullum 2002: 130). That this also applies to Late Modern speech is for instance suggested by Denison (1998: 191)[125] and becomes apparent with a look at the conversational narratives in the OBC, of which (64) may serve as a characteristic example (switches are indicated with horizontal lines ||):

(64) Christmas Eve I was going to see a friend over in the Borough, I was asked to come over and sup with them; I was coming through Bow church-yard, there was another strange young man coming along; || I **see** something laying against the church wall, and he and I **goes** and **looks** at it, || it was a parcel just by the pump; || so **says** I, here is a parcel here, || and he took it up on his back, and carried it a considerable way past Garlick-hill, when I took and carried it till that gentleman stopped me with it.
(OBC, t17950114)

In his attempt to explain how he came into possession of stolen goods, the accused Thomas Clarke switches between present and past tense four times and uses the present-tense forms *see*, *goes*, *looks*, and *says* in a narrative of past events. The label "historic present"[126] is applied to such verbs that are "present tense in form but past time in reference" (Johnstone 1987: 34) and therefore present a structural alternative to past tense forms. Formally, the historic present may make use of forms that show non-

[125] In fact, Denison (1998: 191) states that the historic present was available for use "in appropriate circumstances" throughout the Late Modern period. Although he does not explain what is meant by "appropriate circumstances", it seems reasonable to consider conversational narrative one such "appropriate" context.
[126] There are a number of other terms: narrative present, dramatic present, historical present or conversational historical present (CHP).

standard agreement, as evidenced by *he and I **goes** or **says** I* in (64), where the *-s* inflection is found with subjects other than 3rd person singular. This does not have to be the case, however, as the form *I **see*** in (64) illustrates. As Chapman (1998: 38) notes, the historic present is "a special marked tense", to which "the rules for the distribution of *-s* in subject-verb concord do not apply": for instance, historic present forms with *-s* can also be found with an adjacent, non-coordinated personal pronoun subject (as in *I goes* or *we says*).

Interest in tense switching, whether in narrative in general or in reporting clauses in particular, has usually been focussed on establishing the functions of using an 'unusual' or 'marked' tense form and possibly establishing rules for switches between tense forms. The two major lines of explanation found in the literature either consider the historic present a marker of immediacy or involvement or interpret the alternation between historic present and narrative past as a structuring device for narratives.

In line with the first tradition, most English grammars assert that the historic present's purpose is to make past events more dramatic, vivid or immediate (e.g. Quirk et al. 1985: 181, Biber et al. 1999: 454, Huddleston & Pullum 2002: 130). Its use is often associated with the notion of 'involvement'. Rühlemann (2007: 192), a recent study on English conversation, claims that the historic present serves to mark the speaker's involvement and contributes to the audience's involvement in a narrative. Adopting the concept of empathetic deixis presented in Lyons (1977), Rühlemann (2007: 192) interprets the tense shift from narrative past to historic present as a move from "origo-farther" to "origo-nearer" reference, arguing that "present tense, as reference to present time, would have to be located near the origo". As this shift towards the speaker's origo at the same time involves a shift towards the recipient's origo, the historic present strengthens involvement in the narrative for all parties concerned.

The idea that the historic present conveys immediacy or vividness and, in a way, transports past events into the present moment has often been criticised (e.g. Wolfson 1979, 1982, Johnstone 1987, Fludernik 1991). Wolfson (1979: 169) condemns such accounts as "usually vague and often linked with pseudo-psychological claims as to the state of the narrator's involvement". She further points out that the so-

called English present tense "is not used to refer to present action, except in the sense that it includes the moment of speaking, as when it is used to express general truth or habitual action" (Wolfson 1979: 179). This line of argument casts doubt on interpretations that consider present tense forms in conversational narratives as linked to present-time reference. Instead, Wolfson (1979: 172, 181) argues that the use of the historic present can only be meaningfully interpreted in alternation with the past tense and that switches between the two alternatives structure a narrative by separating different episodes in a story.

While Schiffrin (1981) and Fludernik (1991) largely accept this assessment of tense shifting as a structuring discourse feature, the notion of involvement is not completely abandoned in their work. Fludernik (1991: 392), for instance, argues that verb tense in oral narrative has only a differential but no temporal function, but believes that the historic present can be employed to mark a speaker's emotional involvement in a story. Based on Fleischman's (1990: 55) observation that in narrative, the past form is the unmarked option and a present-tense verb the marked option, she claims that the historic present is used to signal "tellable events" (Fludernik 1991: 392) that help furnish a story with the 'point' that makes it worth telling in the first place (Labov 1972a: 366–375). Similarly, Schiffrin (1981: 59) sees the historic present as one of several 'internal evaluation devices' (along with e.g. the progressive) that contribute to the point of a story.

Whether they see tense shifting primarily as a structuring device or primarily as a way of signalling involvement, all these scholars agree that tense choice in past narratives does not provide information on the temporal sequence of events: Schiffrin (1981: 51) proposes that tense switching is almost exclusively restricted to what Labov (1972) calls the 'complicating action' part of a narrative, which relays a series of temporally ordered events that make up a story. As the temporal order of events is therefore clear, the switches are free to function as an 'internal evaluation device'. This proposed restriction of tense switching to the complicating action cannot be upheld based on data from other studies (e.g. Rühlemann 2007: 191), but I would argue that the temporal sequence of events can indeed be made clear by other means than verb tense, such as adverbs (Wolfson 1979: 180), which frees the historic present to be put to other uses.

6.1.2 Tense shifting in discourse introducers and the form *I says*

As it is "exceedingly common in everyday language" that interlocutors share with each other what was said in earlier conversations (McCarthy 1998: 151), tense shifting can also readily be observed in discourse-introducing verbs such as SAY. There are indications, though, that tense shifting in discourse introducers might function differently than in other environments in past narrative. Wolfson (1979: 178–179) notes that alternation between past and historic present for the verb SAY, which is pervasive in direct speech presentation in her data, does not serve to demarcate turning points in the story or separate different events.

Ultimately, she suggests that the enormous frequency of SAY caused a "loss of significance through overuse" where switches between *said* and *says* are concerned (Wolfson 1979: 178). Johnstone (1987: 42) puts forward a different explanation: discourse introducers are said to follow a different set of rules because "verbs like *say* or *go* do not carry the sort of lexical meaning that other verbs do", but function as "semantically neutral place markers". The function of tense switching in discourse introducers is instead linked to indexing authority or attitude. Johnstone (1987) claims tense shifting in reporting verbs serves to encode the (changing) status relations between the persons in a narrated event and offers narrators the possibility to manipulate their 'footing',[127] i.e. the projected self of the speaker. Focussing on tense shifting in reporting clauses with SAY, Johnstone (1987: 41) finds "authority *says*/nonauthority *said* alternation" and concludes that authority figures' speech is introduced with historic present while non-authorities' words are introduced with a past tense verb. In a similar vein, Sakita (2002: 85) argues that reporting-verb tense expresses "reported-speaker attitudes (or, more precisely, speakers' mental images of reported-speaker attitudes) that are not otherwise dialogue-externally explicit". For instance, the use of the past tense is associated with attitudes of conflict and challenge, while the present tense is associated with excuse and retreat (Sakita 2002: 95).

What is problematic about these accounts is the repeatedly expressed caveat that narrators might make use of tense shifting for the stated purposes, but that their choices are by no means predictable. Sakita (2002: 102) cautions that "tense realization in each story is individual and particular to each situation" and Johnstone (1987: 50)

[127] The term 'footing' is taken from Goffman (1981: 128), where it is defined as "the alignment we take up to ourselves and the others present as expressed in the way we manage the production or reception of an utterance".

emphasizes that storytellers have their "own individual, creative reasons for making the choices they make". Ultimately, this means that where the observed tense shifts align with differences in speaker attitude, status or authority, we have to assume the validity of such an approach. Whenever we find shifts that cannot be linked to authority or attitude issues in the way described above, we still have to assume the validity of the approach, as these deviances are, after all, a product of speakers' creativity. In the end, this has little explanatory power.

Historic present reporting verbs, especially the form *I says*, are also discussed in Rühlemann (2007), a more recent study of English conversation based on the BNC. Historic present *says* is interpreted as typical of the language of conversation: the tense, i.e. the historic present, is considered a strategy for achieving audience involvement (as mentioned above) and the form, i.e. *-s* inflection with any subject, is interpreted as a strategy to alleviate the processing load in real-time conversation. The generalisation[128] of the *-s* inflection to the entire present-tense paradigm of SAY reduces production and processing pressures both on the morphological level (by having only one inflectional ending throughout the paradigm) and the phonological level (by having the same vowel throughout the paradigm) (Rühlemann 2007: 176).

The fact that no instances of *I say* were found in a random sample from the BNC's conversation subcorpus further supports the idea that the *-s* ending is generalised. In fact, *I say* serves different functions in conversation. Mainly, it acts "as a discourse marker referring to present discourse rather than as a preface to presented speech" (Rühlemann 2007: 172–173). The latter is the domain of *I says* and its structural alternative *I said*. These two are, in fact, exclusively used to introduce past speech in past narratives. The absence of *I say* in texts where *I says* occurs frequently as a direct discourse introducer had been noted by other authors before (e.g. Johnstone 1987: 38), but can for the first time be successfully accounted for with Rühlemann's (2007: 172-176) explanation*: I says* is an unambiguous alternative to *I say* and comes with a lower processing load for those involved in conversation, he argues. It is

[128] Rühlemann (2007: 166–167) defines generalisation as the "tendency for a number of forms to be generalised to other grammatical functions, apart from those they carry out in non-conversational registers" and states that this process "concerns particularly verb forms".

frequently found in "point-counterpoint exchanges with rapidly changing turns", where these advantages are particularly useful (Rühlemann 2007: 175).[129]

Use of the historic present in discourse introducers is almost exclusively found in spoken language or when spoken interaction is portrayed in writing. A study of *I says* in the BNC indicates "unambiguously that *I says* is almost entirely restricted to conversation" (Rühlemann 2007: 170). The form *says* seems even to be a sort of stereotype associated with spoken interaction: Fludernik (1991: 392) states that forms like *says he* were employed in literature from the Early Modern period onwards in order to mimic spoken conversations.

Only very little research exists on the social dimension of the historic present in narratives. The already-mentioned case study on *I says* in Rühlemann (2007) is a notable exception, which finds that the use of the reporting clause "depends heavily on the sociological variables of sex, age, class and dialect" (Rühlemann 2007: 178). The BNC shows "some evidence that female speakers use it more often than male speakers" and "clear evidence" that speakers aged between 35 and 44 years, i.e. middle-aged speakers, use it most (Rühlemann 2007: 178). The feature is predominantly found among lower middle class speakers and is "decidedly untypical of upper-class language" (Rühlemann 2007: 178). Its regional distribution in Britain shows that it is "fairly widespread" in the north of England and in Ireland, but only infrequently used elsewhere (Rühlemann 2007: 178).

6.1.3 Late Modern grammars on *I says/I said*

In contemporary English*, I says* is considered a vernacular feature (Rühlemann 2007: 167) because of its marked (nonstandard) inflection. It is widely considered nonstandard and grammatically unacceptable (Carter & McCarthy 2006: 823). In Late Modern English, *I says* – as well as nonstandard concord in general – was heavily stigmatised by contemporary grammarians.

Among the 187 grammars surveyed in Sundby et al. (1991: 138), 38 contain critical remarks on using a 3rd person singular verb with the pronoun *I*, termed '*I* V3' in the book. Although only 20% of all grammars in the survey tackle this specific

[129] In contemporary English, *I goes* fulfils the same functions: it, too, serves as a multi-turn quotative in extended stretches of reported conversation with frequent speaker changes (Rühlemann 2008: 173).

problem (and provide examples of it that they consider incorrect), the issue of subject-verb agreement in general is widely included, yielding over 2,000 examples in 18th-century grammars (Sundby et al. 1991: 103). Among the entries specifically concerned with *I* V3, we find works with a broad temporal (1754-1799) and geographical range (American, English, Scottish and Irish publications), suggesting that the feature was generally considered unacceptable. Most grammars criticise it as bad English, ungrammatical or even absurd, or decry it as a solecism; the label 'colloquial' is applied to it in one source (Sundby et al. 1991: 138). Although the survey of grammars does not list all examples that the grammars branded as incorrect and it is therefore impossible to ascertain whether the verb SAY was mentioned in particular in every source, it stands to reason that all 3SG verb forms with *I* – including those of SAY - were stigmatised.

Only three of the 16 grammars from the 19th century that were surveyed for the present study mention the particular issue of using an *-s* form with a non-third person subject. Where it is mentioned, it is criticised harshly: Crombie (1809: 332) calls constructions like *I reads* a solecism, an "offence against the rules of syntax", Pinnock (1830: 102) criticizes *I sings* as a concord error, and Beard (1854: 97) brands *I does* and similar uses as uneducated and wrong. Furthermore, all 16 surveyed grammars emphasise that there must always be formal agreement between subject and verb. Failing to produce such formal agreement is considered an error in all these publications. Interestingly, we find three comments on the use of the historic present in narratives. Turner (1840: 44) remarks that "[i]n animated narrative the present tense is sometimes substituted (by the figure enallage) for the imperfect". Similar comments are found in Rushton (1869: 189) and Mason (1873: 51). None of the examples given for this phenomenon include nonstandard inflection, though. Use of the verb SAY in narratives is not directly addressed.

6.2 Methodological considerations

The present study is exclusively concerned with one verb (SAY) in one construction (i.e. in combination with the first-person subject pronoun *I*) fulfilling one function (introducing direct discourse). This combination recommends itself for study as it

provides a clear focus, is frequently found and widely exhibits variation. Restricting the study to discourse introducers is also prudent in light of research suggesting that the use of historic present and past tense forms works differently in these contexts than in other parts of conversational narrative (see 6.1).

It has been shown that the historic present in narrative is "especially common with speech act verbs like SAY or GO" (Biber et al. 1999: 455), and that direct discourse presentation[130] is the most frequent way of introducing anterior discourse in spoken conversation (McIntyre et al. 2004: 69, Rühlemann 2007: 123). We can thus expect to find many instances of the variable in the OBC, a corpus that is made up of past narrative to a large extent.

Example (65) shows both inflectional possibilities (*said* and *says*) in discourse introducers in combination with various pronouns and in two syntactic structures (V-S, i.e. *says I* and *says she*, and S-V, i.e. *I said*).

(65) An Beldam. One Morning **I [...] said**, How d'ye do, Mrs. Ray? I can't do well, **says she**, when I have got such a Rogue of a Husband. Her Arm was as black as a Hat, and so was her Thigh, for she took up her Clothes and shew'd me - such an Arm, and such a Thigh, I never saw in my Days! Lauk a dazy! **says I**, what have you married? (OBC, t17340630-15)

That SAY is the reporting verb found (multiple times) in this example is not unusual. In contemporary English, it is one of the major reporting verbs in conversation (Rühlemann 2007: 128), and the combination *I says* has been identified as particularly frequent (Rühlemann 2007: 131, 172). In the Late Modern period, SAY was probably even more important because its strong present-day 'rivals' in that domain, GO and BE LIKE, are more recent innovations.[131]

All combinations of contiguous *I* + *says/said* were extracted from the OBC and the CLMET-drama.[132] This includes the inverted possibilities *said I* and *says I*. In the

[130] Direct discourse presentation (or direct speech) is just one of several possible reporting modes. A major distinction is usually made between five: in addition to direct discourse presentation (*He said, "Sorry, I'm late".*) on one end of the continuum and indirect discourse presentation (*He said he was sorry that he was late*.) on the other, speakers may also opt for one of three intermediary forms (see Leech & Short 2007: ch. 10).

[131] D'Arcy (2017: 23) reports that quotative BE LIKE was introduced by speakers born in the 1960s.

[132] In the OBC, some reporting clauses between 1725 and 1755 are in round brackets, or parentheses, instead of between commas. Consider the following example:

> *I think, Sir, (**says this pretended Lady of mine**) that it's now high time to undeciive [sic] you: - I don't question but that you think you have marry'd a rich Lady of Barbadoes; when, indeed, you are quite mistaken. Mistaken! (**says I in a great Surprize**) Why, pray Madam, what are ye? I am now your Wife, says she; but before you made me so I was Mrs. Eccleton's Maid.* (OBC, t17250827-63)

remainder of this study, any mention of *I says* should be taken to include both *I says* and the inverted option *says I*, just as *I said* is intended to encompass both *I said* and *said I*, unless otherwise specified. The list of results was cleaned of all instances in which *I says* or *I said* did not function as reporting clauses introducing direct discourse presentation, as is the case in example (66).

(66) **I said**, I was sure he was mistaken. (OBC, t18050220-40)

Nonstandard pronoun usage, i.e. *me* instead of *I*, occurs only once in the OBC (67). This example was left out of the analysis.

(67) D– ye for a Son of a Bitch, **says me**, I have got none of your Money, and if I had, what then? (OBC, t17260425-27)

Other forms of SAY (that is, apart from *says* and *said*) almost never occur in discourse introducers: *saith*, while attested (infrequently) in the OBC, is never found in this context. There is only one single use of *I say* as a direct-discourse introducer in past narrative, shown in (68).

(68) Hannah Moses. […] when I found him at home, I asked the prisoner for some halfpence; so he *say* to me do you want any halfpence, **I *say*** yes, he went to the brown bag and gave me sixteen pence for a shilling.
(OBC, t178607190154)

The boldface in this example was added by me, but the italicisation of *say* was in the original printed issue of the *Proceedings*. This suggests that the printers (and potentially even the scribe who created the original transcript) found this usage strange enough to mark it. Hannah Moses was from Amsterdam and, English not being her first language, perhaps did not know that *I says* is conventional in such contexts. In general, though, *I say* is restricted to introducing present discourse, as in (69).

(69) Mr. Silvester. […] Can you, gentlemen, suppose, that men of this description […] should so far forget themselves? but, Gentlemen, **I say**, it is a very common observation, that the Devil will forsake his friends at the last; and, in this case, he certainly has forsaken two of his very best friends.
(OBC, t17870418-118)

The use of round brackets is relatively rare: for SAY + I, 2 examples with *I said*, 6 with *said I* and 18 with *says I* were found. Round brackets were only used in reporting clauses in mid-position in a sentence. Remarks in the *Printer's Grammar*, a printing manual from the 1750s, indicate that parentheses were still used in such a function at the time, but were going out of style and increasingly replaced by commas (Smith 1755: 104).

This functional distribution of *I say* and *I says/said* is in accordance with the findings for contemporary English in Rühlemann (2007).

In the end, all tokens were coded for the factors mentioned in Table 28.

Factor	**Levels**
FORM	Historic present Past tense
STRUCTURE	Pronoun-first (*I says/said*) Verb-first (*says/said I*)
PERIOD	1720-1769 1770-1819 1820-1869 1870-1913

Table 28. Coding for analysis of *I says / I said*

In the OBC examples, coding for SOCIAL CLASS and GENDER was automatically added.

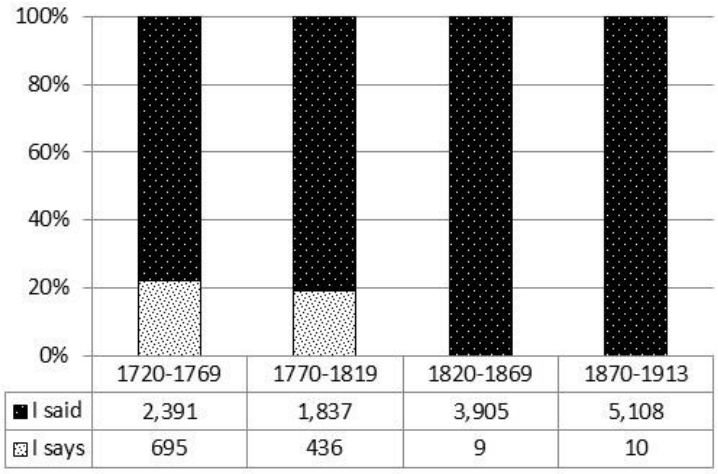
6.3 Preliminary analysis

The results of the corpus study of *I says/said* based on the OBC and the CLMET are presented in this section and the following (6.4). As the analysis of this variable strongly suggested the presence of scribal interference, the structure of this analytical part deviates slightly from those dealing with the other variables: Section 6.3.1 presents the general diachronic development of *I said/I says* in the OBC, focussing on the unexpected trends observed, before Section 6.3.2 puts forward an explanation for these trends, arguing that interference by scribes is key to understanding the OBC developments. Section 6.4 goes on to discuss the results in detail.

6.3.1 *I says/I said* in the OBC: An unexpected picture

In total, 14,391 relevant tokens for *I* + *says/said* were retrieved from the OBC. With 13,241 instances, *said* is clearly the preferred variant, making up 92% of the extracted items. 1,150 tokens of *says* were retrieved. However, the distribution of the variants across time is highly uneven, although the feature is not implicated in change: Figure 40, showing the frequencies and percentages of *I says* and *I said* across four

subperiods, clearly illustrates that *I says* is practically non-existent in the 19th century, but takes up a share of about 20% in both 18th-century periods.



	1720-1769	1770-1819	1820-1869	1870-1913
I said	2,391	1,837	3,905	5,108
I says	695	436	9	10

Figure 40. *I says* and *I said* in reporting clauses by period in the OBC (N = 14,391)

To gain more insight into the development of the variants, a more fine-grained periodisation by decades is employed in Figure 41. This figure not only confirms the overall downward trend in the 18th century and the near-absence of *I says* in the 19th century,[133] but also provides crucial additional information on the presence of *I says* in the OBC.

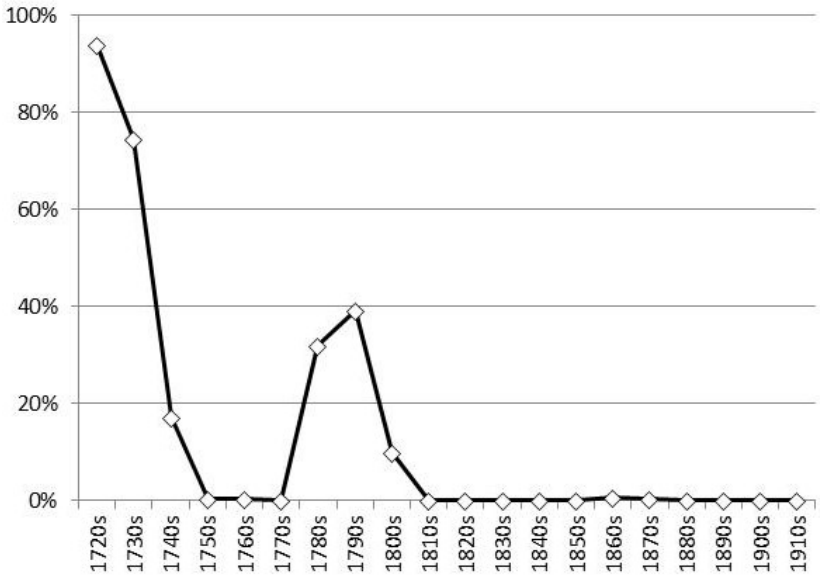


Figure 41. Percentage of *I says* in the OBC, by decade

[133] In fact, after 1800, *I says* never reaches the 1% mark again (for detailed figures, see Appendix: Table A-3).

The decade-by-decade breakdown reveals a remarkable dip in the use of *I says* from the 1750s to the 1770s (percentages range between 0.3 and 0.4%), then a resurgence before the ultimate disappearance of *I says*.

This is a highly unusual development. After all, it is unlikely that a generation of speakers simply did not have a variant in their repertoire that the preceding and following generations employ. Closer inspection of the figures reveals that the development illustrated in Figure 40 is not an artefact of data scarcity for individual decades: with the exception of the 1720s, all decades furnish at least 350 tokens - most of them considerably more. An additional look at *says* in combination with other personal pronouns and with other NPs (see Figure 42) shows that the development is not just restricted to *I says*, but to the use of *says* as a discourse introducer in general:

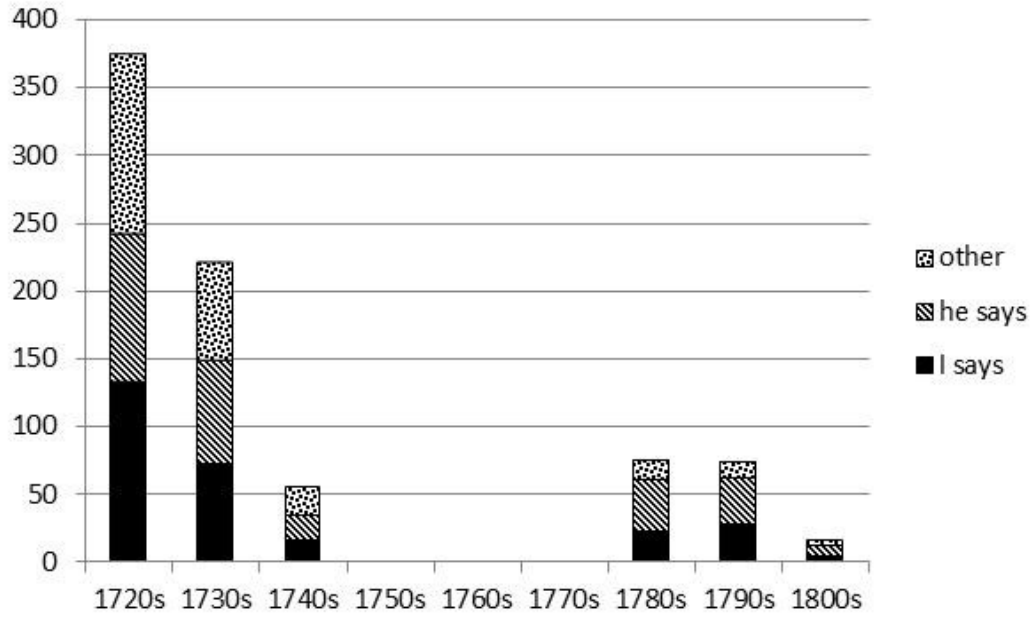


Figure 42. Discourse introducers with *says* p100tw in the OBC between 1720 and 1809: *I says* - *he says* - other (including other pronouns and NPs + *says*) (N = 3,716)

Figure 42 shows the normalised frequencies per 100,000 words[134] of discourse introducers with *says* in the OBC, distinguishing the groups 'I + *says*', 'he + *says*' and 'other' (including *says* + other personal pronouns and + NPs). As only 58 additional instances with *says* are found after 1809, the diagram only considers data between

[134] Normalised frequencies are employed here instead of percentages because extracting and vetting all instances of *said* + pronoun/NP would have been beyond the scope of this study. For absolute figures, see Appendix: Table A-4.

1720 and 1809. It is evident that the pattern *I* + *says*, which makes up roughly a third of all reporting clauses with *says*, is not the only one that shows a decline. Even more importantly, the other patterns equally show (near-) absence in reporting clauses from the 1750s to the 1770s. This internal inconsistency in the corpus merits closer inspection.

In a source like the OBC, which is based on transcripts made in court, it is possible that such strange disruptions are the results of scribal interference. Many different scribes worked on the *Proceedings* throughout their history. That one or several scribes working during the period in question suppressed a variant is much more likely than a hypothetical scenario in which either general language use or a particular register (almost) completely abandoned one of two alternatives in one generation and then reintroduced it. It seems that the downward trend as such could be explained by changing register conventions in the courtroom (increasingly formal), but that the episode of near-absence in the mid-18th century is artificially created at the level of the scribes.[135] This hypothesis is explored in the following section.

6.3.2 The case for scribal interference

A first look at the data revealed that discourse introducers with *says* (including *I says*) basically vanish for a span of roughly 30 years, which is at odds with the overall pattern of development for the variable in the corpus. To determine whether scribal interference is at work here, this section explores variation by scribe, and also assesses the external fit of the OBC results for *says* and *said* (Schneider 2013) so that conclusions on the relation between the depiction of *says/said* in the *Proceedings* and more general developments in language use can be drawn.

It was already noted in 6.3.1 that the trajectory of *(I) says* in the OBC points to a lack of internal consistency for the variable *said/says*. Next to internal consistency, external fit is an important indicator of whether corpus results actually represent linguistic reality or are effects of one of the 'filters' (e.g. scribes) imposed on actual language use (Schneider 2013: 73). The basic idea is thus to check the results of one corpus against those of previous research and against results in other corpora. Similar

[135] I focus on scribes because they were closest to the interaction in the courtroom and thus closest to the original speech event. Unfortunately, we only have incomplete information on who transcribed the *Proceedings* throughout their long history (see Appendix: Table B-1).

results would be an indication that it is indeed extralinguistic reality that is presented in writing, while disparate results would suggest that other factors are at play. Due to a lack of research on this phenomenon in Late Modern English, I limit my comparison to the CLMET only. As the drama subcorpus of the CLMET only yields 58 tokens for the variable (32 for *I said*, 26 for *I says*) throughout the entire period under consideration and is thus not suitable for a detailed analysis, I resorted to the subcorpus of narrative fiction (abbreviated CLMET-narrfic in the following), which represents another speech-related genre.

A comparison of the percentages of *I says* in the CLMET with the corresponding OBC results is shown in 30-year periods[136] in Figure 43. The figure also includes information on the average percentage of *I says* for the period 1720-1839 in both corpora.

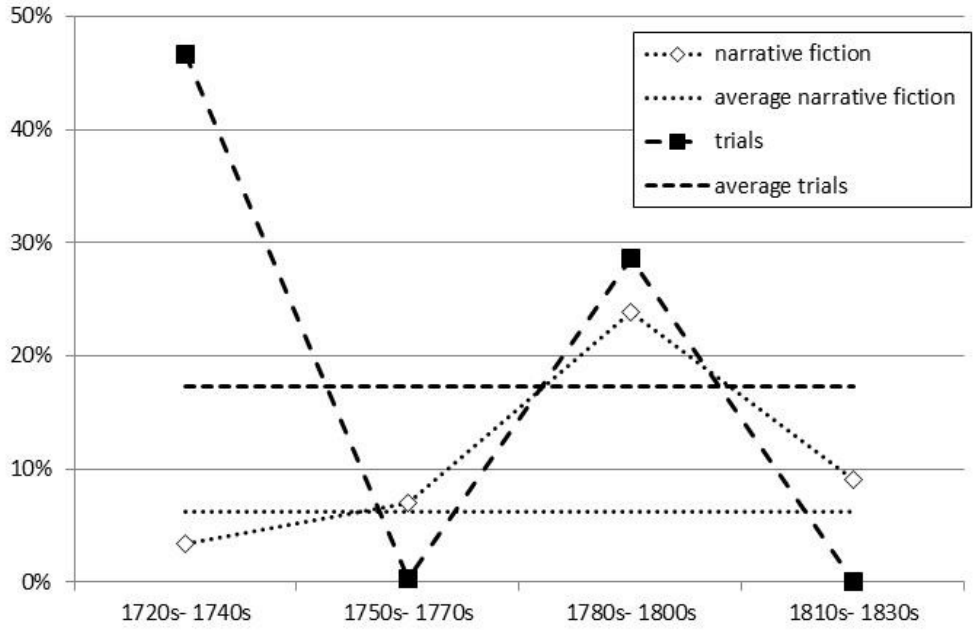


Figure 43. Percentages of *I says* in the OBC and the CLMET-narrfic, plus average percentage of *I says* in the OBC (17.3%) and the CLMET-narrfic (6.2%) between 1720 and 1839

The absolute frequencies can be found in Table 29.

[136] Sorting the CLMET texts into 30-year periods had to be done by hand based on individual texts' years of publication. Instead of using decades as in Figure 42, 30-year-periods were chosen because some decades in the narrative fiction corpus yielded only very few tokens (e.g. 1880s: 4; 1780s: 13).

		I said	*I says*	% of *I says*
1720s-1740s	narr. fiction	1,328	47	3.4%
	trials (OBC)	787	689	46.7%
1750s-1770s	narr. fiction	479	36	7.0%
	trials (OBC)	1,992	7	0.4%
1780s-1800s	narr. fiction	67	21	23.9%
	trials (OBC)	1,082	435	28.7%
1810s-1830s	narr. fiction	200	20	9.1%
	trials (OBC)	1,546	1	0.1%

Table 29. Absolute frequencies of *I says* and *I said* and percentages of *I says* in the OBC and the CLMET-narrfic between 1720 and 1839 (N = 8,737)

It is evident that external fit is a problem here: the sharp drop in the 1750s-1770s in the OBC, a corpus that reports a higher average value of *says* than the CLMET-narrfic, is not found in the CLMET-narrfic, but is apparently unique to the trial proceedings.

Finally, it is useful to take a look at the historical context in which this drop in the *Proceedings* takes place to see if an external reason for the development can be identified. As we know from the grammars of the time, *I says* was considered false concord and heavily condemned (see 6.1.3). It is therefore not far-fetched that it should be removed from official publications. However, it seems curious that this only happened at a specific time (1750s-1770s). A major shift in editing policy took place around 1778, when the *Proceedings* became an official record – this is too late to account for the unusual development of *I says*. Another option is that interference took place at a level much closer to the source, i.e. at the scribal level.

To explore this possibility, variation by scribe is listed in Table 30.

Scribe	**active duty**	***I said***	***I says***	**% of *I says***
Unknown	1720-1736	84	458	84.5%
Gurney, Thomas (?)	1737-1748	689	222	24.4%
Gurney, Thomas	1749-1770	1,712	15	0.9%
Gurney, Joseph	1770-1782	418	1	0.2%
Blanchard, William	1782	57	10	14.9%
Hodgson, E.	1783-1792	301	217	41.9%
Sibly, Manoah	1793-1795	155	120	43.6%
Ramsay, William[137]	1795-1800	150	66	30.6%

Table 30. Variation between *I says / I said*, by scribe, for the period 1720-1800 (information on scribes based on Huber 2007 and Canadine 2016)

[137] Between 1795 and 1797, Mr. Ramsay was transcribing the *Proceedings* together with Mr. Marsay, another scribe.

Two scribes stand out in this overview: Thomas and Joseph Gurney, father and son, are responsible for the lowest ratios of *I says* (both below 1%). Their tenure as scribes coincides with the low figures for this variant: it is clearly documented that they were court scribes between 1749 and 1782, and it is likely that Thomas Gurney also transcribed some earlier trials between 1737 and 1748 (see fn 37, 3.2.3). To account for the latter circumstance, there is a row in the table labelled 'Gurney, Thomas (?)'.

Based on these observations, it seems likely that scribal interference led to *I says* practically vanishing between the 1750s and the 1770s. As both *said* and *says* were represented by different combinations of symbols in shorthand, this cannot be an effect of the transformation of shorthand notes into longhand manuscripts.[138] Rather, I think it was an intentional change. My best guess is that the Gurneys opted not to use *I says* in their transcriptions as a rule, and instead changed it to *I said* because they considered *I says* false concord and therefore inappropriate in written transcripts. I am not suggesting that they interfered with a large number of features, but simply changing an inflection seems to me in the realm of possibility. It is, after all, a relatively minor change, easily made.

To be fair, this hypothesis cannot explain why some very few instances of *I says* remain in the *Proceedings* during the decades in question: a look at Table 30 shows that their number is not 0, but 16, during the Gurneys' tenure. If one looks more closely at the individual tokens of *I says*, it turns out they do not share any common characteristics that could explain their special status. Instead, they are uttered by a number of different speakers (of different genders and social classes), were printed during different printers' tenures,[139] and occur at various positions in the utterances. No generalisations suggest themselves. It is of course possible that these tokens were simply overlooked by the scribes.

If we assume that scribes interfered with this feature, why did *I says* experience a rise after the Gurneys' tenure? The variant certainly had not become more acceptable by then. It is possible that the resurgence of *says* is connected to the City of London's demand in 1778 that any proceeding should contain a "true, fair and perfect narrative" of the trials – this could have encouraged scribes like Blanchard, Hodgson and Sibly to

[138] Thomas Gurney wrote a book explaining his shorthand system (*Brachgraphy)*, which yields information on how verbal inflection could be expressed in shorthand (Gurney 1752: 7, 15-19).
[139] For a list of printers that were active when the Gurneys acted as scribes, see Appendix B: Table B-1.

be more accurate and actually write down what they heard. Increasing regulation could thus have actually had a positive impact on linguistic faithfulness in this case. That *I says* and *I said* alternate in the 18th century proceedings – apart from the Gurneys' work – seems to indicate the *Proceedings'* faithfulness to original speech for this feature. Neither *I says* nor *I said* are a default option that scribes always use to introduce speech. In the 19th century, however, the situation had changed: *I says* must be assumed to have vanished not only from the records of spoken courtroom interaction, but also from the speech of trial participants, as it had become relegated to informal language.

At any rate, the preceding discussion leads me to believe that the factor SCRIBE should be added to the analysis for this particular linguistic feature, which will be done in the following. As the preliminary results clearly indicate that there is practically no variation between *says* and *said* in the 19th century in the OBC, the feature will only be investigated in detail in the 18th-century *Proceedings.* The comparison between the OBC and the CLMET will include data from both the 18th and 19th centuries, though.

6.4 Findings and discussion

After a preliminary look at the distribution of *I says* and *I said* in the OBC in 6.3, the present chapter provides a more detailed analysis of the findings. The OBC results are discussed in 6.4.1, and Section 6.4.2 compares findings in the OBC and the CLMET throughout the Late Modern period. Section 6.4.3 offers some concluding remarks.

6.4.1 *I says/I said* in the OBC: results

Since scribes may have actively shaped the use of the feature *I says/I said* in the OBC, the coding scheme outlined in Table 28 (featuring the independent variables STRUCTURE and PERIOD) was amended to include the variable SCRIBE (see Table 30 for levels of this factor). As always, the OBC results are additionally coded for GENDER and SOCIAL CLASS. In accordance with the procedure outlined in 3.5, a logistic regression model for the variable *says/said* was created.

In the end, only the factors STRUCTURE and SCRIBE turned out to significantly impact the distribution of *I says* and *I said*, as shown in Table 31. The second column contains the log odds of *I says*, i.e. indicates the likelihood of this variant being used.

	estimate *b*	**SE**	**z value**	**p-value**	**confidence intervals**	
					2.5%	**97.5%**
(Intercept)	-2.2309	0.2658	-8.394	<0.001	-2.7671791	-1.72412564
STRUCTURE = Verb-first	5.6089	0.2251	24.917	<0.001	5.1844111	6.06861120
SCRIBE= Gurneys	6.3029	-0.3411	-18.477	<0.001	-7.0037424	-5.66301847
SCRIBE= Gurneys (?)	-3.3666	0.2446	-13.765	<0.001	3.8639740	-2.90422604
SCRIBE= later	-0.4845	0.2646	-1.831	<0.01	-0.9990355	0.04064667
Concordance Index *C*		0.82				

Table 31. Output of logistic regression including predictors STRUCTURE and SCRIBE; based on OBC

The values indicate that the STRUCTURE verb-first increases the odds of *I says* occurring compared to pronoun-first, and that *I says* is most strongly disfavoured when the Gurneys – either Thomas or Joseph – act as scribes. For the variable SCRIBE, it is necessary to explain that several levels were conflated as the differences between them were not significant: out of the initially eight distinctions made in Table 30, only four remain: 'earlier' (i.e. the unnamed pre-Gurney scribe(s)), 'Gurneys(?)' (i.e. the proceedings between 1737-1748 which were potentially transcribed by Thomas Gurney); 'Gurneys' (including transcripts by Thomas and Joseph Gurney, which did not differ significantly with regard to use of *says/said*) and finally 'later' (post-Gurney scribes, which did not differ significantly from each other, either).

The graphic representations of the effects of STRUCTURE (Figure 44) and SCRIBE (Figure 45) illustrate the impact of these variables.

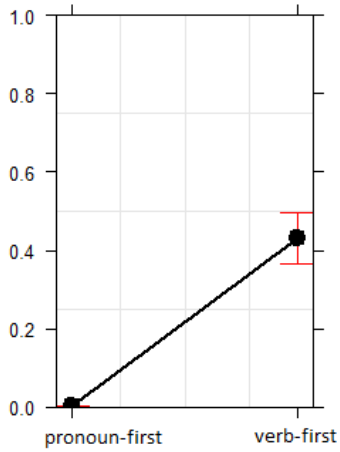


Figure 44. Effect of factor STRUCTURE on likelihood of *I says* (data: OBC)

Figure 44 shows that neither verb-first nor pronoun-first constructions are *says*-favouring in the sense that it would be a majority option in these – the overall proportion of *I says* in the data is much too low for that. However, it becomes evident that verb-first constructions do represent a far more hospitable environment for *I says* than pronoun-first constructions. While the chance of *I says* in verb-first constructions is estimated at 43%, a probability of less than 1% is calculated for *says* in pronoun-first structures.

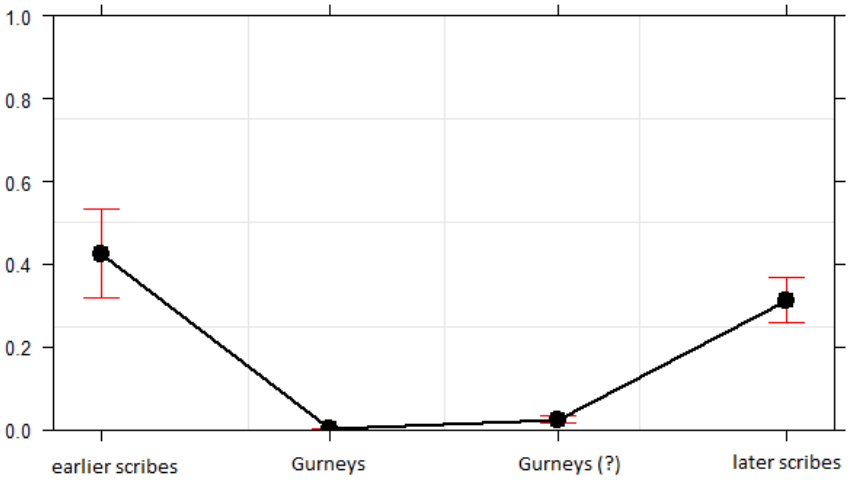
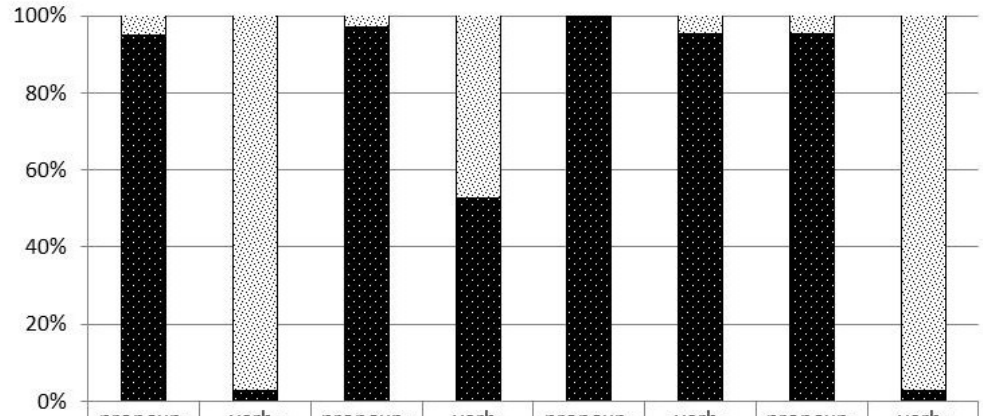


Figure 45. Effect of factor SCRIBE on likelihood of *I says* (data: OBC)

As for the scribes, the chances of *I says* occurring are computed to be nearly 0 (below 0.1%) for the Gurneys. In the period when it was unclear whether Thomas Gurney was

already in charge ('Gurneys(?)'), the probability of *I says* occurring is 2.5%, also very low. Among earlier and later scribes, the chances of *I says* are much higher in comparison: in the texts taken down by (an) unknown previous scribe(s), the chance of *I says* is calculated to be 42.3%, for those following the Gurneys, the probability is 31.1%. Particularly the fact that a higher degree of variation is found with the 'later' scribes strengthens the argument for scribal interference.

The diachronic development and the interplay of the two factors STRUCTURE and SCRIBE can be illustrated with a look at the observed frequencies from the corpus for the 18th century. Figure 46 provides an overview of raw frequencies and percentages.



	pronoun-first	verb-first	pronoun-first	verb-first	pronoun-first	verb-first	pronoun-first	verb-first
	earlier: unknown scribe(s) (1720-1736)		Gurneys (?) (1737-1748)		Gurneys (1749-1782)		later: Blanchard, Hodgson, Sibly, Ramsay (1782-1800)	
I says	4	454	15	207	1	15	32	381
I said	72	12	458	231	1,838	292	653	10

Figure 46. *I says* and *I said* by SCRIBE and STRUCTURE, OBC (N = 4,675)

The verb-first option is always more likely to include *says* than the alternative, irrespective of scribe. With the miniscule numbers of *I says* for the Gurneys (1749-1782), this does obviously not amount to a significant difference. Nevertheless, *says* seems to be very much associated with this construction in 18th-century trials. It should be interesting to see whether this also holds for other Late Modern texts.

6.4.2 *I says/I said* in comparison to the CLMET

A comparison to the narrative fiction subcorpus of the CLMET contextualizes the findings from the OBC. A regression model with the predictors CORPUS (OBC – CLMET), STRUCTURE and PERIOD was run. As Table 32 indicates, all predictors significantly influence the distribution of *I says* and *I said*. The estimates refer to the probability of *I says* occurring.

	estimate *b*	SE	z value	p-value	confidence intervals	
					2.5%	97.5%
(Intercept)	-9.0890	0.2027	-44.847	<0.001	-9.4946813	-8.6999054
STRUCTURE= Verb-first	6.0251	0.1467	41.078	<0.001	5.7429909	6.3181755
CORPUS= OBC	3.3612	0.1034	32.511	<0.001	3.1617997	3.5672690
PERIOD== 1780-1850	1.4755	0.1060	13.914	<0.001	1.2699554	1.6858788
PERIOD= 1850-1913	0.5035	0.1616	3.116	<0.01	0.1828497	0.8170221
Concordance Index *C*		0.82				

Table 32. Output of logistic regression including predictors STRUCTURE, CORPUS and PERIOD; based on OBC and CLMET-narrfic

There are small but significant effects of the predictors CORPUS and PERIOD: the chances of *I says* are never high, but texts from the OBC and texts from the period 1780-1850 are most likely to contain the variant (see Figure 47 and Figure 48).

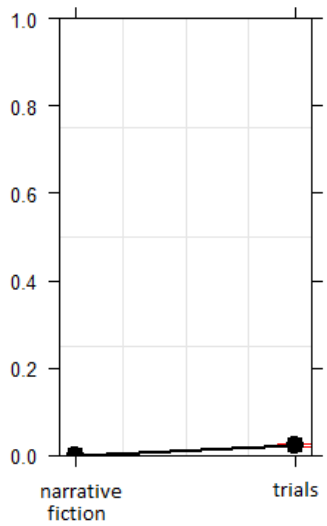


Figure 47. Effect of factor CORPUS on likelihood of *I says* (data: OBC and CLMET-narrfic)

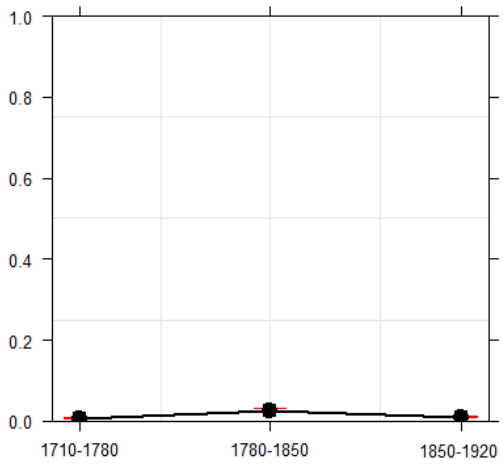


Figure 48. Effect of factor PERIOD on likelihood of *I says* (data: OBC and CLMET-narrfic)

As the diagrams show, 'most likely' in this case still amounts to very unlikely: The chance of *I says* occurring is only at 2.3% in the OBC (compared to below 0.1% for the CLMET) and at 2.6% for the period 1780-1850 (compared to below 1% in the other periods). The effect of the predictor STRUCTURE, which was the strongest effect when looking at the OBC in isolation, is much more substantial, as Figure 49 illustrates.

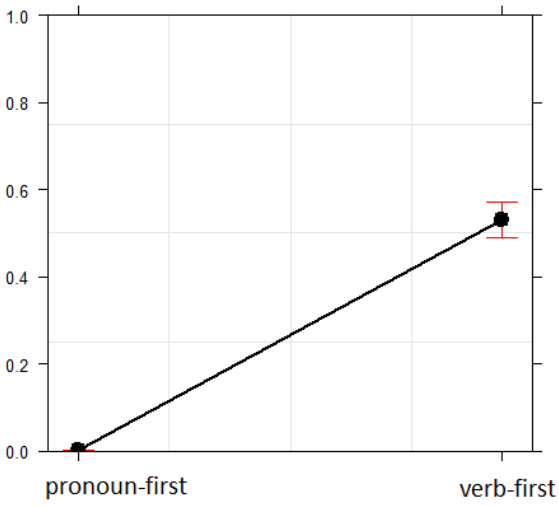


Figure 49. Effect of factor STRUCTURE on likelihood of *I says* (data: OBC and CLMET-narrfic)

Encountering *says* in pronoun-first contexts is very unlikely (below 0.1%), but the chance rises dramatically in verb-first constructions: the model predicts 53% of such constructions to contain *says*.

A look at the observed frequencies in both corpora can add more depth here. Table 33 provides absolute figures and percentages of *I says* by corpus and period.

		I said	***I says***	**% of *I says***
1710-1780	narrative fiction	1,807	83	4.4%
	trials (OBC)	2,779	696	20.0%
1780-1850	narrative fiction	687	59	7.9%
	trials (OBC)	3,455	436	11.2%
1850-1920	narrative fiction	1,637	53	3.1%
	trials (OBC)	7,007	18	0.3%

Table 33. *I says* and *I said* in the CLMET-narrfic and OBC, by period (N = 18,718)

Figure 50 graphically depicts the proportional development of the variant *I says* in the two corpora. Note that the periodisation here is based on the built-in periodisation of the CLMET.

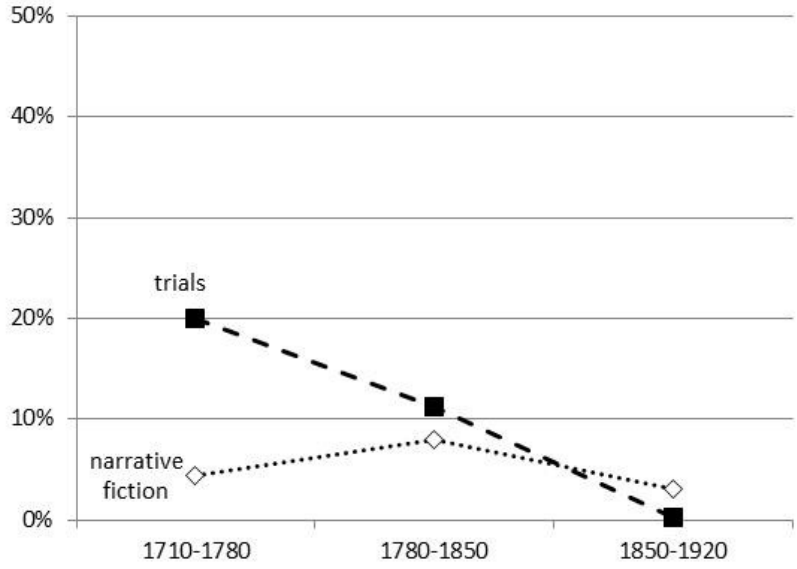
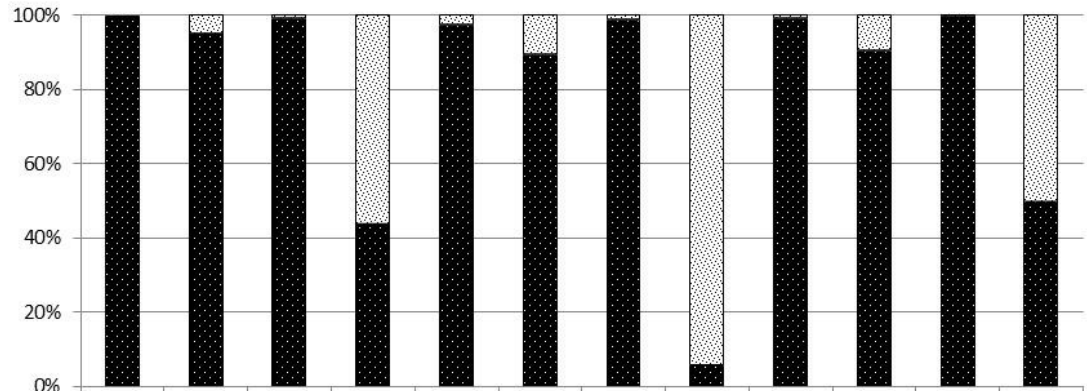


Figure 50. Proportion of *I says*, by period, in the CLMET-narrfic and the OBC

There is a clear downward trend in the OBC, but a more or less stable, if small, proportion of *I says* (between 3.1 and 7.9%) in the narrative fiction corpus. The trial proceedings are originally more open to the variant *I says* (which represents 20% of all tokens in the first period) – even though the almost complete suppression of *I says* in the 1750s, '60s and '70s is included in the figures for the first period. However, this changes with time. In the final period, the proportion of *I says* in the OBC drops to only 0.3% – and thus below the proportion recorded for the narrative fiction corpus, which is at 3.1%.

A look at the impact of the verb-pronoun structure throughout time is provided in Figure 51: it is interesting to note that in all periods, verb-first constructions (I + SAY) show a greater proportion of *says* than verb-second constructions (SAY + I). This is true in both corpora, although the difference between the constructions is more pronounced in the OBC data than in the CLMET.



	1710-1780 narrative fiction pronoun-first	1710-1780 narrative fiction verb-first	1710-1780 trials pronoun-first	1710-1780 trials verb-first	1780-1850 narrative fiction pronoun-first	1780-1850 narrative fiction verb-first	1780-1850 trials pronoun-first	1780-1850 trials verb-first	1850-1920 narrative fiction pronoun-first	1850-1920 narrative fiction verb-first	1850-1920 trials pronoun-first	1850-1920 trials verb-first
I says	0	83	20	676	6	53	40	396	11	42	17	1
I said	156	1,651	2,245	534	236	451	3,429	26	1,236	401	7,006	1

Figure 51. *says* and *said* by structure, by corpus, by period (N = 18,717)

It is worth looking at these numbers from another perspective, namely in terms of the choice of construction (verb-first vs. pronoun-first) with SAY as such, independently of the form of SAY. Table 34 shows the percentage of verb-first constructions by corpus and period:

		pronoun-first	verb-first	% of verb-first
1710-1780	narrative fiction	156	1,734	91.7%
	trials (OBC)	2,265	1,210	34.8%
1780-1850	narrative fiction	242	504	67.6%
	trials (OBC)	3,469	422	10.8%
1850-1920	narrative fiction	1,247	443	26.2%
	trials (OBC)	7,023	2	0.0%

Table 34. Constructions with SAY by period and corpus (N = 18,717)

The figures indicate that the type of construction that is strongly associated with the variant *says*, i.e. the verb-first variant, is gradually disappearing from the texts. In the narrative texts, it was the clear majority option at the beginning of the Late Modern period (91.7%), but then dropped to 67.6% and finally 26.2% in subsequent periods. Never that high to begin with in the trials, the proportion also diminished there: 34.8% became 10.8% and finally close to 0% in the following periods. With the decline of the

verb-first construction, the context in which *says* is the favoured option disappears over time.

The issue of scribal interference is an added complication in the OBC, so to speak. To put this problem in perspective, an additional comparison across corpora may be useful. Table 35 compares the figures for *I says* and *I said* in the narrative fiction section of the CLMET, the OBC, and a version of the OBC that excludes all texts created by the Gurneys, called 'OBC -Gurneys'. Thanks to the OBC's size, there is still a sufficient number of tokens in the latter corpus to allow for a useful comparison of the three corpora.

		I said	***I says***	**% of *I says***
1710-1780	CLMET-narrfic	1807	83	4.4%
	OBC	2779	696	20.0%
	OBC -Gurneys	774	680	46.8%
1780-1850	CLMET-narrfic	687	59	7.9%
	OBC	3455	436	11.2%
	OBC -Gurneys	3331	436	11.6%
1850-1920	CLMET-narrfic	1637	53	3.1%
	OBC	7007	18	0.3%
	OBC -Gurneys	7007	18	0.3%

Table 35. *I said* and *I says* in the CLMET-narrfic, the OBC, and the 'OBC -Gurneys', by period

The periods in which the absence of texts by the Gurneys influences the figures are shaded in grey in the table: instead of values of 20% for the first and 11.2% for the second period, we record 46.8% for 1710-1780 and 11.6% for 1780-1850. Scribal interference thus leads to two effects: it shrinks the proportion of *I says*, especially in period 1 (the differences between the proportions are relatively small for period 2 because the Gurneys were only active until 1782), and it leads to an underestimation of the overall downward trend for *I says* throughout time (because it underreports use in earlier periods). In the 'OBC -Gurneys' corpus, a plunge of more than 45 percentage points is observed between periods 1 and 3 (instead of a reduction of 20% in the OBC).

6.4.3 Conclusions

For the variable *I says/said*, different patterns emerge in the two corpora: relative stability with a very low rate of *I says* in the CLMET and a clear downward trend in

the OBC. An added complication is presented by some OBC scribes who severely underuse *I says*. Common ground is found in the importance of constructions (verb-first or pronoun-first) for the choice of variant in both corpora.

The general decline noted in the OBC may well be a side-effect of the continuing professionalisation and formalisation of trials and trial proceedings. Trials grew increasingly structured and technical, with a greater extent of work falling to legal professionals. All this presumably increased linguistic self-consciousness in the courtroom. In addition, court proceeding grew more and more formalised. It makes sense that such developments negatively impact a variant like *I says*, which was considered informal and evaluated as incorrect by contemporary grammarians. Apart from the Gurneys' tenure as scribes between 1750s and the 1770s, when *I says* was actively suppressed, the variant thus gradually declined in the OBC based on speakers' stylistic choices on what was appropriate courtroom register.

The OBC and the CLMET both provide evidence for the importance of a linguistic factor in the distribution of *says* and *said*: the linguistic structure in which SAY occurs heavily influences the choice of form. Inverted constructions (SAY + I) drastically improve the likelihood of *says*. Interestingly, these inverted constructions go out of use with time in both corpora, which fuels the decline of *says*. They manage to retain a foothold for longer in the CLMET narrative fiction corpus, though, where they make up about a third of all discourse introducers with I + SAY between 1780-1850. This could be due to the invented nature of narrative dialogue and the use of inverted discourse introducers (often with historic present) as a sort of literary imitation of speech. Fludernik (1991: 392) argues that *says he/she* (a pattern closely related to *says I*) with past reference was a literary "standard pattern already in Shakespeare and Defoe, since it mimetically recreates what is believed to be the colloquial standard" (see 6.1.2). That the proportion of inverted constructions between 1780-1850 had already dropped to 10.8% in the OBC could actually be an indication of greater faithfulness to actual spoken usage in the *Proceedings* – at least as far as the structure is concerned. Based on the non-literary OBC texts, the actual colloquial standard in the Late Modern period was probably the pronoun-first structure. This pattern is also by far the most frequent one in present-day English: the entire BNC yields only 5 tokens of *says I*, as opposed to 1,163 instances of *I says*.

An association of the variants with different social factors could not be established. Gender and class cannot be shown to significantly impact speakers' choices for this feature. The regional distribution of *I says*, which is reported to be strongly restricted to northern dialects in contemporary English (Rühlemann 2007: 178), could not be tested systematically, as the relevant information for speakers/authors is not available in the corpora used. However, there is no evidence of a restriction to northern areas: after all, the variant is present in the OBC, which features mainly speakers living in southern England. At its largest, the court's jurisdiction included London, the County of Middlesex, metropolitan Essex, Kent and Surrey. It may well be that in Northern areas, *says* was even more frequently used at the time, and that usage there remained stronger while it later diminished in the south. Further studies with regionally diverse data are needed to explore this issue.

6.5 Summary

This chapter investigated variation in the tense of the discourse-presenting verb SAY, to be more precise whether the verb is used in the past tense or the 'historic' present tense form (*I said – I says*). After setting the scene with a general introduction to tense shifting in narrative and in discourse introducers (6.1) and highlighting some methodological issues (6.2), the results based on the OBC and the CLMET were reported and discussed in the previous Sections 6.3 and 6.4.

It is now time to re-evaluate the hypotheses developed in 3.6, which predicted stable variation between *I says* and *I said*, with different proportions for the variants depending on corpus/genre, and the stigmatised variant *I says* being rare in trial transcripts due to the formality of the situation and the rather formal genre conventions. Analysis of the data makes it plain that we need to review these predictions. While it is true that stable variation with a rather low proportion of *I says* occurs in the CLMET, the OBC shows a different pattern: the proportion of *I says* is initially much higher than in the CLMET (contrary to the assumption that it would be rare in all trial proceedings), but declines dramatically so that its proportion in the OBC is actually lower than in the CLMET at the end of the Late Modern period. This is interpreted as an effect of more formalised trials, which went hand in hand with

increased attention paid to language by speakers, leading them to discard informal forms. In addition, the form *says* is also adversely affected by the decline of the inverted verb-pronoun pattern (*says I*), with which it is strongly associated.

A major difficulty is presented by an unusual drop in the proportion of *I says* in the OBC in the 1750s, 1760s and 1770s: the form is almost not found. After establishing that this pattern is not recorded for other genres and is not likely to be the result of speakers' language choices, a case was made for scribal interference leading to this effect. Obviously, this anomalous pattern in a part of the corpus also represents a methodological challenge: it is, for instance, much more difficult to realistically assess the impact of speaker parameters like gender or class. A feature like *says*, which is "so clearly 'against the rules' of Standard English" (Rühlemann 2007: 169) and harshly criticised in Late Modern grammars, normally would be a prime candidate for social differentiation (which is reported for contemporary English), but no effect is found in the OBC. However, as speakers' choices are probably not reported faithfully for this feature, this might simply be an effect of the material.

The size of the OBC turned out to be a major benefit for the study of this feature, though: in spite of almost complete suppression of *I says* between the 1750s and the 1770s, the OBC yields a much higher relative frequency of the feature than the other Late Modern corpora I consulted: the drama subsection of CLMET yields 12.4 tokens pmw, the narrative fiction subcorpus 19.1 tokens pmw, while the OBC boasts 82.2 tokens pmw. This clearly illustrates the value of a corpus of transcribed speech for tracing features like *I says*, an informal discourse management phenomenon. The following chapter is also devoted to a conversational phenomenon which is widespread in contemporary spoken English: past tense BE variability, more precisely variation between *you was* and *you were*.

7 Subject-verb agreement: *you was – you were*

> Did not the woman of the house say, that **you was** in such a state of drukeness [sic] that **you were** incapable of going home, and the servant should see you home?
> (OBC, t17860719-37)

Variability of the verb BE in terms of concord with the subject, especially in the past tense, is widely reported across varieties of English. The use of so-called default singulars (e.g. *was* in contexts were standard English today prescribes *were*) is even considered to be a vernacular universal, and has been noted both synchronically and diachronically. The present chapter is concerned with a special case of past BE variability, namely variable agreement of past tense BE with the singular subject pronoun *you* in Late Modern English.

In contrast to other types of BE variability such as default singulars with existential *there*, variation between *you was* and *you were* is "a late innovation in the language" (Nevalainen 2009: 99–100). In the Late Modern period, singular *you was* emerged along with singular *you were*, presumably in combination with the reorganisation of the second-person plural paradigm, i.e. *thou* disappearing as a singular pronoun and *you* branching out into singular uses.

Previous studies find a marked increase of *you was* in the first half of the 18th century and record a decline of this form after the mid-century mark. Some scholars assume that language prescription favouring *you were* played at least a part in the decline of *you was*, turning it from a variant used by all social classes to a sociolinguistic stereotype, excluded from the standard and associated with lower-class or dialectal usage (see e.g. Laitinen 2009: 200).

The present chapter first reviews previous research on BE variability, with a focus on *you was/you were* (7.1), before discussing important methodological issues (7.2) and presenting the findings based on the OBC and CLMET data (7.3). Section 7.4 provides a summary of the chapter.

7.1 Previous research and treatment in LModE grammars

Contemporary standard English prescribes the use of *was* with first and third person singular, and that of *were* with second person singular and first, second, and third person plural. In spite of this, a great deal of variation is found in usage, especially in more informal situations, giving rise to e.g. *you was* instead of *you were*. The present section first outlines BE variability in general (7.1.1), taking note of different levelling tendencies found in contemporary varieties as well as variability in a diachronic perspective. Sections 7.1.2 and 7.1.3 are concerned with *you was/you were* only: previous research on variation in Late Modern English and contemporary grammarians' comments on the phenomenon are reviewed, respectively.

7.1.1 The larger context: variable agreement patterns of BE

The variable considered in the present study, i.e. singular *you was/were*, is just one of the many kinds of variation that the verb BE routinely exhibits across varieties of English. Another particularly prominent example is variation with *there* existentials and notional plural subjects in both past and present tense (*there is/are penguins – there was/were dogs*).

In present-day English – and arguably also in Late Modern English – the verb BE is special insofar that it is has many distinct forms showing person and number agreement with different subjects in the standard variety (see e.g. Schilling-Estes & Wolfram 1994: 275). In contrast to this, English verbal morphology in general is "pervasively regular", which makes distinct tense variants minority forms and linguistically marked (Hay & Schreier 2004: 210). As such marked variants often face pressures toward analogy, especially in vernacular varieties, it is no surprise that the verbal paradigm of BE often shows levelling tendencies meant to "bring irregular person-number concord in line with the vast majority of verbal paradigms that display no such agreement" (Hay & Schreier 2004: 210).[140]

Past tense BE is especially prone to regularisation (see Hay & Schreier 2004, Chambers 2009): Britain (2002: 17) remarks that all vernacular varieties of English

[140] Even accounts that dispute the existence of person and number agreement of verbs with their subjects in contemporary English in general make an exception for BE, which represents a special case: Hudson (1999: 173), for instance, argues that in contemporary English, "person is irrelevant to all verbs except BE, and [...] past-tense verbs and modals (other than BE) have no number agreement features".

seem to have variable past tense BE, "even those varieties with relatively little other morpho-syntactic non-standardness". In particular, extension of the form *was* is very widespread: so-called "default singulars" with plural subjects (as in *Diana and I was the last ones*) even make Chambers' (2009: 258) list of "vernacular roots" or "vernacular primitives" that recur in vernacular dialects across the globe.[141]

However, it is not necessarily the form *was* that serves as the pivot form in levelling. In fact, three broad tendencies are reported: 1) generalisation to *was*, 2) generalisation to *were*, 3) a mixed system ordered by polarity: *was* for positive polarity, *weren't* for negative polarity (Anderwald 2001: 9–10; also see Britain 2002: 17–19, Moore 2010: 347). Pattern 1 (generalisation to *was*) is the "most common" pattern (Britain 2002: 17–19). It can also be considered the "most predictable" one, as the form *was* is much more frequent than *were* even in the standard and thus presents a logical choice of levelling pivot (Anderwald 2001: 9).[142] The use of only past tense *was* neutralizes the singular-plural distinction of the system and thus aligns it with other past tense paradigms (Anderwald 2001: 9). The same effect is achieved by applying Pattern 2 (levelling to *were*), which is less frequently found (Anderwald 2001: 9). The third option, Pattern 3, is a mixed system, combining generalised *was* in positive clauses and generalised *weren't* in negative clauses (Anderwald 2001: 9).[143] This levelled pattern, with morphological distinctions based on polarity rather than on person and number, arguably makes more cognitive sense than the standard pattern and is more in line with cross-linguistic principles (Anderwald 2001: 17–19).

The importance of polarity for past BE variation is confirmed in many other studies, e.g. Britain (2002), reporting on the English Fens, or Cheshire & Fox (2009), researching London usage. However, the constraints are not always the same: in the Scottish town Buckie, Smith & Tagliamonte (1998: 118) found that negative polarity categorically required *was* in the local variety.[144] Smith & Tagliamonte (1998: 118–119) argue that this finding, which is "dramatically different" from findings on *was/were* use in other communities, "suggests that negative constructions are indeed implicated in whatever process underlies *was/were* variation more generally".

---

[141] In the first edition of the book, Chambers (1995: 242) identified *we/you/they was* as a vernacular root.
[142] For more factors that make *was* a 'predictable' choice of levelling pivot, see Schilling-Estes & Wolfram (1994: 276) or Britain (2002: 37).
[143] For further references on studies detailing these various patterns, see e.g. the overview in Moore (2010: 347).
[144] In affirmative contexts, *was* is also the dominant option in Buckie: it occurs 72-91% of the time (Smith & Tagliamonte 1998: 118).

Diachronically, extensive variation throughout the paradigm of BE has been documented:[145] alternation among distinct patterns in Old English, particularly in existentials, is described in Quirk & Wrenn (1958), and the development of BE irregularity in the Middle and Early Modern periods is traced in e.g. Jespersen (1961), Traugott (1972), Visser (1963-1973) and Denison (1998). For the 16th to 18th centuries, Tagliamonte (2009: 104) notes "rampant" variation based on information in the 1989 edition of the OED (*be*, v.; Simpson & Weiner 1989). Studies like Laitinen (2009) and Nevalainen (2009) attest to continued variation in the Late Modern period.

It should be noted that systems of concord with BE are historically variable. Preferences for (potentially levelled) systems such as these described above may change over time: in London English around 2000, older speakers mainly showed *were*-generalisation whereas younger speakers preferred a mixed system, which may indicate a change underway (Anderwald 2001: 14). The development of BE with plural NP subjects in New Zealand English over the past 150 years makes a strong case for the influence of extralinguistic factors on this process and demonstrates "the possibilities of nonlinearity in language change" (Hay & Schreier 2004: 233): from the dialect contact phase of the creation of New Zealand English, which was accompanied by pressures towards standardisation, singular concord in both existential and nonexistential environments (e.g. *the girls was outside*) declined. By 1900, singular concord in nonexistentials had practically disappeared. However, singular concord in existentials (*there was stars in the sky*), which had always been present to a greater degree, began to increase in the 20th century.

According to Hay & Schreier (2004: 233), this reversal of the trajectory of change was possible because existentials became "dissociated from the nonexistentials", which "liberated [them] from the standardizing force" and allowed the increase of singular concord in existentials, which shows high rates in modern New Zealand English. Default singulars in existential constructions have also reportedly been on the rise in many other contemporary varieties (Tagliamonte 2009). Britain & Sudbury (2002: 210) even suggest that the rise of *was* with following plural NPs in existentials represent "a change presently underway in most (all, even the standard?)

[145] This section draws on the overview of diachronic studies presented in Hay & Schreier (2004: 210). For further references to historical studies, see the discussion in Tagliamonte (1998: 153–157).

varieties of English". Colloquialisation may be aiding the rise in singular agreement that is observed for the recent past (Collins 2012: 60).

Previous research has also come to the conclusion that levelling is determined by universal principles, but at the same time subject to local constraints: based on a comparison of 13 varieties of English, Tagliamonte (2009: 114) concludes that default singulars, for instance, show no "specific universal (vernacular) hierarchy according to grammatical person" (like *was* always being more frequent with *you* than *they*, for instance). Instead, these hierarchies differ according to local constraints from variety to variety. However, they are subject to "fairly consistent scale-independent contrasts *within* the grammatical person hierarchy" (Tagliamonte 2009: 116), namely differences between existentials and nonexistentials and structures with pronouns vs. structures with full NPs.

The clearest differences are found "between existential constructions and everything else" (Tagliamonte 2009: 116). The overwhelming majority of the varieties investigated in Tagliamonte (2009) have significantly more default singulars (i.e. *was*) in existentials than in other contexts where *were* would be required in standard English. Historically, existential *there* constructions were also the most typical environments for default singulars since the late Middle English period (see e.g. Martínez-Insua & Pérez Guerra 2006, Nevalainen 2006, 2009). The special status of existentials is also pointed out in research on contemporary English. Martínez-Insua & Pérez Guerra (2006: 191) comment on the "idiomatized character of *there* plus *be*", and Breivik & Martínez-Insua (2008: 358–359) argue that the sequence of existential *there* + singular BE should be regarded as a unit that has grammaticalised to a "presentative signal" indicating to addressees that new information follows.[146] In earlier work, Cheshire (1999: 137–139) had already suggested that existential *there* + singular BE is best described as an unanalysed whole that serves as a device for topic management and in turn-taking, allowing speakers to take the floor in conversations. The structure is assumed to be "stored and accessed as a prefabricated phrase, rather than as a structure that is generated anew each time that it is used" (Cheshire 1999: 138). Just like similar structures in other languages, like French *il y a* or German *es*

[146] This is especially true of contracted *there's*, which other studies have also pointed out: Crawford (2005: 58), for instance, maintains that the combination of existential subject and copular verb should best be considered a formulaic sequence, and sees indications that "the construction is not analysed in the same way as traditional subject-verb agreement structures and even existential constructions with verbs other than *be"*.

gibt, it does not have concord with the following (notional) subject (Cheshire 1999: 139). The different nature of *there* existentials even leads Walker (2007: 148) to warn that "including existentials in studies of other forms of subject-verb agreement is both methodologically and theoretically unsound".[147]

Another "fairly regular contrast" can be established between noun phrases (NPs) + *was* and pronouns with *was* (Tagliamonte 2009: 116). The so-called 'Northern subject rule', e.g. discussed in Ihalainen (1994) and Klemola (1996), needs to be mentioned in this context: according to this rule, which applies beyond past-tense BE, *s*-forms (such as *was*) can be expected to be more frequent after full NPs or when a clause separates subject and verb, than after pronouns (see e.g. Britain 2002: 19–20, Tagliamonte 2009: 114). To what extent this rule can be said to affect (changes in) Late Modern and contemporary English is not easy to establish. As Pietsch (2005: 149–150) remarks, *was/were* were not originally within the scope of the Northern subject rule. However, Anderwald (2001: 11) deems it "possible that remnants of [the Northern Subject Rule] play a role in the north [of England] for such a frequent verb as *be*". There are also indications that the feature was more widespread historically and also extended further into the south of England: some southern sources show 'northern' *-s* with plural subjects, including *was* with plural subjects, in the 16th and 17th centuries already (Kytö et al. 2011: 235, Visser 1963-1973: 72). Thus, the Northern Subject Rule may also have lingering effects today in varieties beyond the north (Anderwald 2001: 11).

In addition to linguistic factors like polarity, linguistic structure (*there* existentials vs. other) or type of subject (NP, pronoun), social factors like class, ethnicity, gender or formality/situation have also been found to impact the distribution of different forms of BE. In contemporary spoken American English, Riordan (2007: 261) notes "strong effects of social and discourse factors" on (non)concord in existentials, in that increasing age and formality of discourse promote concord. In contemporary New Zealand English, the highest rates of singular concord in existentials are found among nonprofessional speakers and men (Hay & Schreier 2004: 233). The importance of different social factors varies, though: Tagliamonte (2009),

[147] Walker (2007) references a personal communication by Stephen Levey as well as work by Cornips & Corrigan (2005) and Wilson & Henry (1998) as the basis for this assessment.

for instance, finds the effect of gender on choosing *was* in standard-*were* contexts to be moderately strong in some varieties, but mostly weak.[148]

An important difference between spoken and written usage and an impact of formality on BE variability has been noted, too. In general, non-standard concord is especially frequent in contemporary spoken material (see e.g. Breivik & Martínez-Insua 2008: 351). In a study on the use of singular agreement after *there* existentials, Cheshire (1999: 137–138) suggests that lively, quick conversation favours prefabricated *there* + singular BE, while more formal speech styles, where time constraints and pressure to hold or gain the floor are less pronounced (e.g. because speaking turns are more routinely distributed), would be more susceptible to standard concord – especially for speakers who have been exposed to prescriptive norms of subject-verb concord. Crawford (2005: 58–59), though, found much singular agreement with *there* existentials in academic lectures, and argues that "the cognitive burden of spoken language outweighs the formality aspect of academic lectures", which ultimately "results in the formulaic use of contracted existential *there* + *be* (*there's*) without conscious reference to the prescriptive rule of agreement". Meechan & Foley (1994: 82), too, are convinced that "non-concord is the norm" in *there* existentials and that educated speakers' exposure to grammatical rules taught in school simply obscures this fact.[149]

In the end, it is clear that a number of linguistic and extralinguistic factors have an impact on BE variability in general. To gain a better understanding of the particular construction in focus here, the following section will outline previous research on singular *you was* and *you were* in Late Modern English.

7.1.2 Previous work on singular *you was – you were*

Variation between singular *you was* and *you were* represents one of the ways in which the variability of BE, outlined above, manifests. While BE variability in general is much discussed in synchronic studies, there is in fact little diachronic work on the issue and

[148] See Walker (2007: 152) for a concise summary of social factors influencing singular agreement (i.e. the choice of a singular form of BE where plural would be prescribed in standard English) in existentials.
[149] Meechan & Foley (1994: 82) also make an interesting point on linguistic analyses: linguistic analyses are written by highly-educated people who have been exposed to these rules about concord, and that, as a consequence, concord is often assumed in most structural analyses – whether that is a realistic assumption is debatable.

even fewer studies of *you was/were* in the Late Modern period. The only exceptions known to me are Tieken-Boon van Ostade (2002) and Laitinen (2009), whose findings this section draws on heavily.

As outlined above, BE variability as such, including past BE alternation (*was/were*), has been established in the historical record from the earliest stages of the Old English period onwards (see e.g. Smith & Tagliamonte 1998: 107, see 7.1.1). However, variation between *you was* and *you were* is only documented to any noticeable extent beginning in the 17th century. In a study of default singulars in Early Modern letters, Nevalainen (2006: 360) only finds "a couple of instances" of *you was* in the CEEC, quoting one from 1661, and reports first attestations in the 1630s in the Chadwyck-Healey Literature Online database, an enormous collection of text. In the Michigan Early Modern English Materials, including text from l500 to 1700, *you was* is only found from the 1650s onwards, and even then represents a minority variant (Nevalainen 2006: 360).

That the emergence of *you was* in the 17th century takes place at a time when *you*, formerly a plural pronoun only, replaced *thou* as a singular pronoun, is deemed no coincidence: instead, the constraints on the use of default singulars changed with time and "interacted with other linguistic subsystems undergoing change, such as personal pronouns" (Nevalainen 2006: 367). Tieken-Boon van Ostade (2002: 96) assumes that the developments of *you was* and *you were* were both part of "the tail end of the process by which *thou* as the singular pronoun was replaced by *you*". The forms *you was* and *you were* both rose in frequency among largely the same trajectory (Tieken-Boon van Ostade 2002: 96). *You was* apparently emerged because speakers felt it was "logical" to use the verb in the singular with a singular pronoun (Laitinen 2009: 207). It is likely that *you was* served as a 'bridge phenomenon',[150] "facilitating the functional spread of *you were* to include singular reference"- at least as far as standard English is concerned (Tieken-Boon van Ostade 2002: 100).[151]

In her analysis of all novels in the Chadwick-Healey Eighteenth-Century Fiction Full-text Database, Tieken-Boon van Ostade (2002) shows that *you was* is still

[150] Tieken-Boon van Ostade (2002) takes over the term 'bridge phenomenon' from Ukaji (1992).

[151] Some sources claim that there was a distinction between second person singular *was* and second person plural *were* from the late 16th to the 18th century, which was lost from standard English afterwards (e.g. Denison 1998: 317, quoting Phillipps 1970: 159). However, this differentiation of the forms *was* and *were* into singular and plural is not what we find in the historical record: *you were* is also found with singular reference, as e.g. the analysis in chapter 7.3 shows.

rare in the first decades of the 18th century, but experiences a rise in the 1740s and peaks in usage in the 1750s. From the 1760s onwards, use of the construction drops again (Tieken-Boon van Ostade 2002: 95). Similar trends are shown in Laitinen (2009), a study of *you was* and *you were* in correspondence between 1681 and 1800: *you was* is present as a minority variant in the earliest decades (proportion of *was* in 1681-1699: 15%), but quickly gains ground so that it represents the majority option in 1720-1739 (63%) and 1740-1759 (56%). Afterwards, the proportion of *was* begins to drop, and sinks to 37% for 1780-1800, the last subperiod under investigation (Laitinen 2009: 206).

As the rise of *you was* occurs slightly earlier in the letters than in the novels analysed by Tieken-Boon van Ostade (2002), it can be assumed that *you was* first appeared on informal genres like personal correspondence, from which it spread to more formal writing (Laitinen 2009: 206). Another study based on correspondence proposes that *you was* was a marker of a coterie style, i.e. "a feature used by and among a small, intimate group of correspondents" (Fitzmaurice 2004: 379) and a "trace of oral communication" that only managed to seep into the most familiar and intimate kinds of writing and served as a marker of intimacy between correspondents (Fitzmaurice 2004: 380). Fitzmaurice (2004: 364–365) proposes that especially people who found themselves "outside the boundaries of the grammar school and university conventions" promoted the use of *you was*: she names Alexander Pope, a Catholic and thus barred from attending university, and Mary Wortley Montagu, who educated herself via reading, as two such people who were more inclined to experiment with linguistic forms.

Social factors are also part of Laitinen's study: he contrasts women's usage with men's, and considers usage among language professionals as opposed to that of other letter writers. His data suggest that *you was* was a male-led innovation (Laitinen 2009: 209) but that men were also the first to abandon the form again (Laitinen 2009: 210). Language professionals lag behind the control group of non-professionals in terms of adopting *you was*, but they are faster than the control group in shifting to *you were* (Laitinen 2009: 214–215). Based on the behaviour of the language professionals and the fact that *you was* emerged earlier in letters than in novels, Laitinen (2009: 206,

214) argues that it represents a change from below in Labovian terms. Its decline, though, happens above the level of consciousness.

In fact, the decline of *you was* in the historical record[152] reflects its exclusion from standard English, where *you were* was adopted as the single 'correct' form. Scholars have repeatedly drawn attention to the fact that *you was* was heavily criticised in contemporary grammars (see 7.1.3 for more details on grammarians' point of view). Langer & Nesse (2012: 613) cite *you was/you were* as an example of purism in standardisation, which renders structures that have been excluded from the standard language "invisible even though they are still a part of the set of constructions used by native speakers". Tieken-Boon van Ostade (2009: 100) argues that *you was* had been "demoted to non-standard usage" by the end of the 18th century due to normative grammarians' influence. It thus dropped out of use in printed works, but continued on in non-standard usage until the present day (also see Tagliamonte 1998: 184–185). According to Laitinen (2009: 215), the language professionals' head start in the shift to *were* is due to their having access to and being aware of proscribed forms via their occupations.

The influence of prescriptive grammars in this development should not be overstated, though. Nevalainen (2009: 99–100), who shows that default singulars in existential constructions also declined throughout the 18th century, stresses that "the supralocal decline in the use of default singulars in general did not begin with eighteenth-century prescriptive grammars", but that they were "nevertheless influential in stigmatising subject-verb nonagreement, thus making certain concord patterns a conscious choice for the educated". Laitinen (2009: 215) argues that the short-lived expansion of *you was* at the beginning of the 18th century represents a change in progress that was interrupted in the standard language when it was in the mid-range stage. Rydén (1984: 514) also speaks of the "arrest of levelling tendencies" in this case.

For the 19th century, there is comparatively little information on the further development of *you was*. Quoting a personal communication by Tony Fairman, Laitinen (2009: 208) states that the form *you was* remained dominant in partly-schooled writers' letters in the early 19th century. In 19th-century Australian society,

[152] Fitzmaurice (2004: 363) calls *you was* as an "ephemeral expression" due to its short-lived acceptance as an alternative to *you were*.

levelling to *was* is also reported to be associated with the language of the lower classes (Fritz 2007: 187, also reported in Smitterberg 2012: 960). By that time, *you was* was already associated with the usage of the less educated and relegated to nonstandard English. In conversational settings and in dialectal use, second person *you was* is still frequently found today (see Tagliamonte 1998: 157 and references therein for *you was* in dialects of English). Its present-day existence leads Anderwald (2017: 288) to assert that variability between *you was* and *you were* must have been present in the 19th century as well.

In present-day English, verbal concord in English is "a potential social marker" and indicative of a person's knowledge of the standard variety (Crawford 2005: 35, referencing Cheshire 1999). As previous research points out, this has its origin in the Late Modern period, where some forms (like *you were*) were arbitrarily elevated to prestige markers while others (*you was*) were excluded from the standard,[153] forming a system where these forms are "used to maintain social distinctions governed by both overt and covert prestige" (Laitinen 2009: 200). One instrument in the codification and standardisation process, grammar-writing, is discussed in the next section.

7.1.3 Late Modern grammars on *you was/you were*

On a very general level, the treatment of *you was* in Late Modern grammars is aptly summarised in Anderwald (2017):

> [...] it is clear that *you was* changed from being actively recommended (for the singular) to being stigmatized over the century, and this stigmatization increases and becomes categorical towards the end of the nineteenth century. (Anderwald 2017: 289–290)

The present section will add more detail to this brief statement.

The 18th-century grammars surveyed in Sundby et al. (1991) show that concord in general was an issue that raised much comment in the second half of the century

[153] It is sometimes assumed that the use of *was* in *there* existentials with plural postverbal components was not (as) stigmatised in the 17th and 18th centuries, as opposed to *was* with singular pronoun *you*. Tagliamonte (1998: 185), for instance, proposes that *there* + nonstandard *was* actually increased over time, aided by this lack of stigma in the Late Modern period. Other studies come to different conclusions, though: Nevalainen (2009) and Nevalainen (2015) report a decline in singular agreement in such *there* existentials in 18th century letters, which is first observed with male writers. This is attributed to subject-verb non-concord being "heavily stigmatized by 18th-century normative grammars" and male writers having easier access to education and thus to prestigious forms (Nevalainen 2015: n.p.).

(see also 6.1.3). The section labelled '*you* + V3' contains all comments on the use of *you* in combination with forms inflected for 3rd person singular, which includes *you was*. A total of 50 grammars published between 1750 and 1800 in England, Scotland, Ireland and North America contain critical remarks on '*you* + V3': the construction is considered, among other things, ungrammatical, improper, a solecism, inaccurate or barbarous (Sundby et al. 1991: 156). Only one single grammar of their selection argues in favour of *you was*: Murry (1778: 23) criticizes *you were* as incorrect when it refers to one person, and advises to use *you was* (Sundby et al. 1991: 156). Anderwald (2017: 286) also reports that some 18th-century grammarians considered *you was* "a legitimate and functionally useful form", sometimes in connection with establishing a singular-plural distinction for the second person.

Interestingly, Robert Lowth, one of the personalities most strongly associated with language prescription in the second half of the century, also used *you was* when it was popular, i.e. when it peaked in the early 18th century, in his private letters to his wife (Tieken-Boon van Ostade 2002: 95). This illustrates that at least some writers made a difference between formal and informal writing. In his grammar, which is supposed to set out guidelines for formal writing, Lowth (1762: 48) harshly criticizes *you was* as an "an enormous Solecism" (also see Görlach 2001: 101, Tieken-Boon van Ostade 2002: 89–90). His informal norm, as seen in his letters, allowed for variation, though (Tieken-Boon van Ostade 2002: 90).

As with all previous features, I analysed the treatment of *you was* and *you were* in 16 selected 19th-century grammars in order to complement the 18th century information given in Sundby et al. (1991). Tieken-Boon van Ostade (2002: 99), which includes some impressionistic remarks on *you was/were* in individual grammars in the Late Modern period, points out that Webster's *Philosophical Grammar* (1807), an American publication, considered *you was* acceptable because the form was used by authors of the time.[154] However, Tieken-Boon van Ostade (2002: 100) also made it clear that a more systematic review of 19th century grammars is called for because "it is during the nineteenth century that the standard/nonstandard distinction will have

[154] As pointed out in fn 114 in chapter 5.1.3, grammarians as a group did neither consistently apply nor share criteria of acceptability. That acclaimed authors commonly used a feature may have held weight for Webster, but did not for all grammarians. Lowth (1762), for instance, was well aware of the widespread use of *you was* in literary works in the 18th century, but did not consider this circumstance an argument in favour of the variant (see Tieken-Boon van Ostade 2002: 99).

made itself felt with respect to the acceptability of *you was*". My survey of 16 British grammars is a first step in this regard.

Most importantly, none of the British grammars I analysed share Webster's viewpoint, i,e. accept *you was*. Instead, two tendencies are found: either the issue is not addressed at all, or *you was* is condemned. The first tendency represents the majority: Half of the grammars (eight out of 16) do not address which form to use with singular *you* at all, because only *thou* is acknowledged as a singular pronoun (also see Anderwald 2017). These grammars only contain verb tables of BE including *thou wast* for the past singular and *you were* for the plural, but do not discuss the issue further.[155] The other eight grammars, which do include some discussion in addition to conjugation tables for BE, all contain more or less explicit condemnation of *you was*. Pinnock (1830: 141) represents the most restrictive example in that he argues against both the use of *you* as a singular form and its combination with singular verb forms:

> Again: "You was in earnest, and you sought attention;"
> should be, "you were in earnest, and you sought attention**."
> Obs. The use of the word you is indefensible;
> but whether it signifies a singular or a plural number,
> the verb must always be in the plural.
> (Pinnock 1830: 141)

Other grammars acknowledge that *you* is also available as a singular pronoun, but are adamant that it cannot combine with singular verb forms like *was*. Crombie (1809: 240–241) makes it clear that "[y]ou is plural, whether it refer to only one individual, or to more; and ought therefore to be joined with a plural Verb". He considers *you was* a solecism and advocates *you were* instead (Crombie 1809: 376). Four further grammars, which acknowledge that *you* may also refer to the second person singular, argue for its use with plural forms of verbs and thus against *you was* (Turner 1840: 133, Bullen & Heycock 1853: 44, Hiley 1853: 53, Higginson 1864: 48). More indirect condemnation of *you was* is found in Allen (1824: 46, 98) and Turner (1840: 169), which both include sentences with *you was* in their exercises on 'false syntax'. Anderwald (2017: 289) also reports the frequent inclusion of *you was* in sentences to be corrected or in examples of incorrect concord in the 19th century.

[155] Two further remarks are warranted here: Bullen & Heycock (1853: 46) also accept *thou wasest* in addition to *thou wast*. Perhaps more interestingly, Mason (1873), which is among the grammars that only recognise *thou* as a singular pronoun, contains some instances of *you were* with singular reference in the explanatory text or examples unrelated to this particular phenomenon, e.g. *Tell me how old you were when your father died* (190).

7.2 Methodological considerations

The linguistic variable constructed for the current study of agreement patterns consists of the choice between singular and plural past tense forms of the verb BE accompanied by the singular subject pronoun *you*. Those represented paradigmatic alternatives at least in the 18th, and perhaps also in the 19th century (Laitinen 2009: 200). Whenever reference is made to 'standard agreement' in the following, the emerging late 18th-century norm, i.e. *you were*, is meant. Any mention of 'nonstandard agreement' refers to *you was*.

In order to retrieve all relevant instances, the OBC and the CLMET-drama were searched for all contiguous combinations of *you* with either *were* or *was*, including negated instances such as (70) and subjunctive uses as in (71) and (72).

(70) I thought **you was** not doing wrong, as it was your mother's house. (OBC, t18161204-33)

(71) Lord MACGRINNON. Why, if **you was** a woman yourself you could not plead better for them than you do. (CLMET3_0_2_92)

(72) DEARTH. […] The awful thing about a son is that never, never—at least, from the day he goes to school—can you tell him that you rather like him. […]
MARGARET. But if **you were** a mother, Dad, I daresay he would let you do it. (CLMET3_0_3_287)

As the focus is on singular contexts only, all instances in which *you* represented a plural pronoun were excluded. Constructions with the pronoun *thou* are theoretically still possible in the early 18th century, but *thou* does not occur with past BE in the OBC, so it need not be taken into account as another alternative.[156] This might be because the OBC materials mostly document usage in London and its surroundings, and *thou* had become rare in London and the east after 1600 (Kytö et al. 2011: 233). Inverted forms, mainly found in questions (*was you? were you?*), were excluded because their distribution in the OBC is heavily skewed towards the groups of lawyers and judges. Other participants very rarely make use of them because they almost never ask questions.

Both *you was* and *you were* are represented from the beginning of the OBC, though: The first occurrence of *was* in combination with a second-person singular

[156] *Thou* as such only occurs 24 times on the OBC.

subject pronoun can be traced to the year 1721 when a witness testified that Thomas Hill, accused of killing another man, said the following:

(73) D---n you, are you not Dead yet? I told your Wife **you was**, and thought I had kill'd you. (OBC, t17211011-42)

A look at the entire collection of digitised *Proceedings* (going back to 1674) reveals no earlier examples, but does attest to the presence of the form *were* in such contexts for the first time in 1692:

(74) **You were** talking to Mrs. Mary Sheriff. (Proc-16920406)

However, *you were* is only used a total of 8 times in all proceedings before 1720. This is most likely a consequence of the *Proceedings* containing very little direct speech presentation before the 1720s.

After retrieving all relevant instances of *you was* and *you were* from the OBC and the CLMET, they were coded for POLARITY (positive and negative), a factor repeatedly shown to play an important role on the distribution of the variants (see e.g. Smith & Tagliamonte 1998: 115, Anderwald 2001: 5). Additionally, information on when the texts were created is provided via the factor PERIOD, as indicated in Table 36.

Factor	**Levels**
FORM	*you was* *you were*
POLARITY	positive negative
PERIOD	1720-1769 1770-1819 1820-1869 1870-1913

Table 36. Coding for analysis of *you was / you were*

The OBC data were also automatically coded for the social factors GENDER and CLASS. This answers the call for the integration of the sociolinguistic background of speakers in studies of *was/were* made in Tieken-Boon van Ostade (2002: 100).

7.3 Findings and discussion

This section presents the results of the analysis of *you was/were* variation in the OBC and the CLMET-drama. Sections 7.3.1 and 7.3.2 provide overviews of the diachronic developments in both corpora. Section 7.3.3 offers some concluding remarks.

7.3.1 *you was*/*you were* in the OBC

The OBC yields 1,314 instances of *you was* and 2,459 instances of y*ou were*, i.e. 3,773 relevant tokens overall. A logistic regression model was fitted based on the principles outlined in 3.5: the most appropriate model for the present variable involves only one predictor, time (divided into four periods, as in previous chapters). This model is summarised in Table 37. The estimates refer to the probabilities of *you were*.

	estimate *b*	**SE**	**z value**	**p-value**	**confidence intervals**	
					2.5%	**97.5%**
Intercept	-1.00420	0.09608	-10.451	<0.001	-1.1954199	-0.8184809
Period= 1770-1819	0.89059	0.10845	8.212	<0.001	0.6800411	1.1054100
Period= 1820-1869	3.45718	0.16443	21.026	<0.001	3.1418501	3.7870730
Period= 1870-1913	5.45155	0.33227	16.407	<0.001	4.8524847	6.1683990
Concordance Index *C*		0.83				

Table 37. Output of logistic regression including predictor PERIOD; based on OBC

POLARITY, which was also coded for, turned out to be without significant impact on the distribution of *you was* and *you were*: the data indicate that there was no systematic functional differentiation based on polarity (in terms of Pattern 3 outlined in 7.1.1) in the Late Modern period. Neither were social factors (GENDER and CLASS) found to significantly impact the distribution of variants.

The only variable that shows a significant influence on the use of *you was* and *you were* in the OBC is time, represented by the factor PERIOD. In line with earlier research, the OBC results show that *you was* was popular in the 18th century, but vanished almost completely in the 19th century when *you were* became dominant. The calculated probabilities of *you were*, which are depicted in the graph in Figure 52, clearly display this.

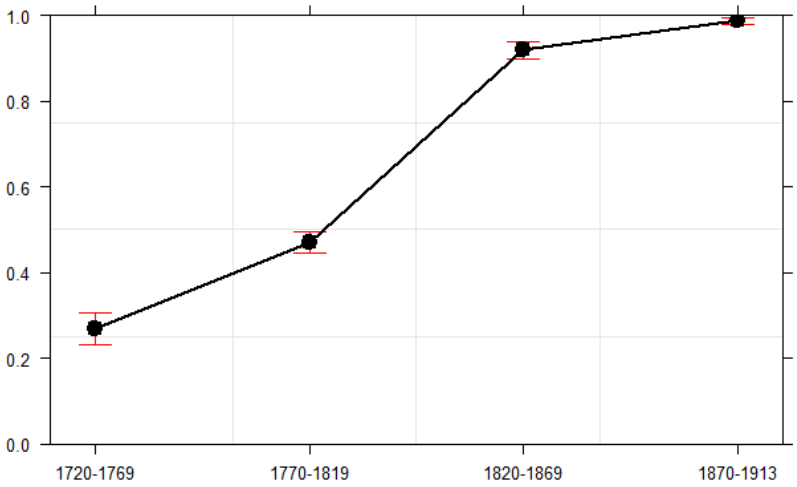
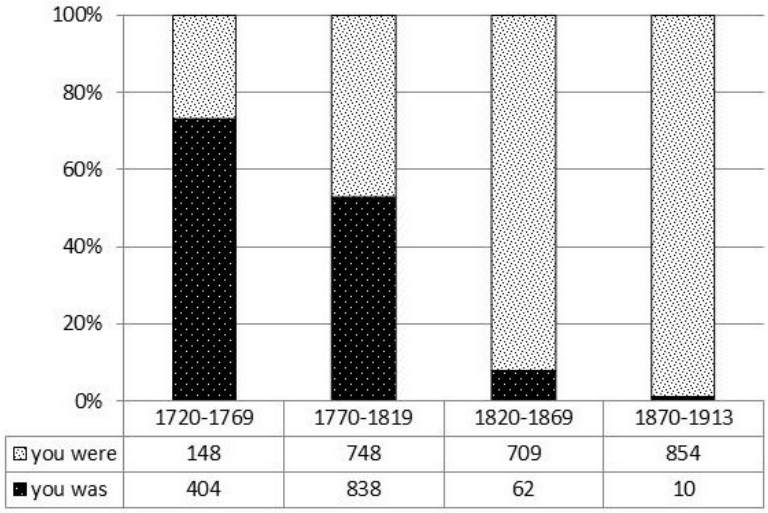


Figure 52. Effect of factor PERIOD on likelihood of *you were* (data: OBC)

While the probabilities of *you were* are below 50% in the 18th century, the 19th century is characterised by a clear predominance of *you were*. The line created based on the model is reminiscent of an S-curve.

If we take a closer look at the observed frequencies through time (see Figure 53), it is evident that the proportions of *you was* in the 19th century are extremely small: the variant makes up for less than 10% in period 1820-1869 and only for a little more than 1% in period 1870-1913.

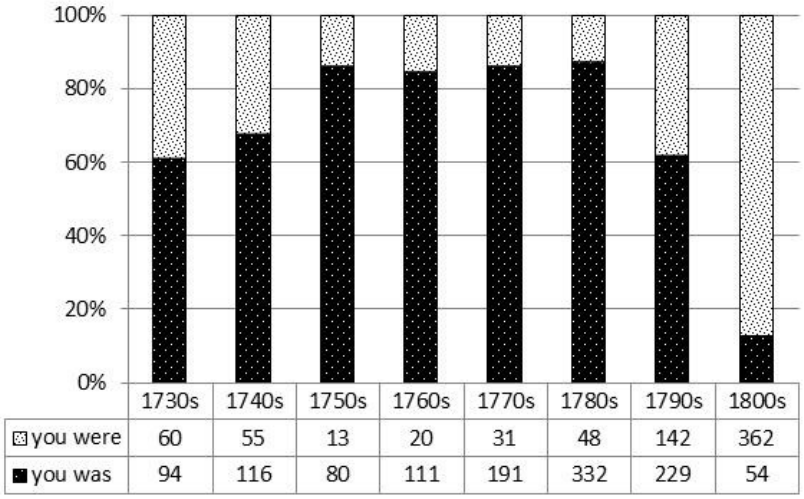


	1720-1769	1770-1819	1820-1869	1870-1913
you were	148	748	709	854
you was	404	838	62	10

Figure 53. *you was* and *you were* in the OBC, by period (N = 3,773)

Although this picture is less extreme than the almost complete disappearance that was recorded of *I says* in the 19th century (see 6.4.1), there certainly are similarities between these two informal variants in the 19th-century trials.

Figure 54 focuses on the development of *you was* and *you were* in the OBC in the 18th century only, when variation was greatest: *you was* represented the preferred choice between the 1730s and the 1790s in the trial proceedings.



	1730s	1740s	1750s	1760s	1770s	1780s	1790s	1800s
you were	60	55	13	20	31	48	142	362
you was	94	116	80	111	191	332	229	54

Figure 54. *you was* and *you were* in the OBC, by decade (1730s-1790s) (N = 1,522)

In other genres, the preference for *you was* did not last as long, as previous research shows. In fictional texts, use of *you was* peaked in the 1750s and started to drop from the 1760s onwards (Tieken-Boon van Ostade 2002: 95). In personal letters, *you was* represented the majority option between the 1720s and the 1750s, and finally becomes a minority choice after 1780 (Laitinen 2009: 206). In the OBC, a drop can only be observed from the 1790s onwards. In comparison to letters and drama, trial proceedings were more hospitable to the variant *you was* over a longer period of time.[157] Though *you was* has been criticised almost from its first appearance, a downward trend in the *Proceedings* occurs only at the end of the 18th century. The variant had been excluded from the formal register appropriate for the courtroom by then. It must be assumed to have remained in everyday conversation – only that we

[157] While the three studies use different set-ups (and periodisation) and comparison therefore has to be taken with a grain of salt, they nevertheless showcase genre-specific trends.

have no record of such informal talk in the OBC. After all, the variant *you was* is still widely found in contemporary conversations.

There are no indications that GENDER or SOCIAL CLASS significantly affect the use of *you was/were*. Certainly, there is not sufficient evidence in the *Proceedings* to confirm Laitinen's (2009) conclusion that men led first in the rise of *you was* and later also in the shift to *you were*. If we break down the OBC figures by gender and period, the distribution in Table 38 emerges.

		you was	***you were***	**% of *you was***
1720-1769	**women**	27	12	69.2%
	men	365	131	73.6%
1770-1819	**women**	10	5	66.7%
	men	819	742	52.5%
1820-1869	**women**	8	71	10.1%
	men	53	627	7.8%
1870-1913	**women**	1	104	1.0%
	men	9	746	1.2%

Table 38. *You was* and *you were* by gender and period, OBC (N = 3,730)

While it is true that the percentage of *you was* is higher among men than among women in the first period (73.6% vs. 69.2%) and that this is reversed in the second period (men: 52.5%, women: 66.7%), the dearth of data points for women (only 39 respectively 15 tokens uttered by women in periods 1 and 2!) makes it impossible to provide a robust assessment of women's linguistic preferences in that period.

While there certainly could be a gender effect, it is not possible to confirm it based on the OBC results. It should be mentioned at this point that the data on which Laitinen (2009) bases his conclusion about a gender effect are not without problems: he relies on very few tokens within a narrow time frame for his assessment, i.e. "the critical period of 1760-1779, i.e., when YOU WAS had started to decline" (Laitinen 2009: 211). For this period, he reports a proportion of 50% for *you was* among women (5 of 10 tokens) and a proportion of 25% among men (28 of 64 tokens).[158] While Laitinen speaks of a statistically significant difference at $p<0.05$ between men and women in 1760-1779, I am unable to confirm this using a chi-squared test, which he

[158] As the overall numbers of tokens in a category were not stated in Laitinen's paper, I calculated them myself based on the figures for *you was* and the proportions given for these figures.

presumably used based on other tests in his paper. So the supposed gender effect is supported by little evidence after all. The data from the trial proceedings in the present study rather point to *you was* as a feature that was equally used by all social groups – probably because variation in spoken language simply was more acceptable for everyone.

7.3.2 *you was/you were* in comparison to the CLMET

Running a logistic regression on the combined OBC and CLMET data confirms the importance of genre, represented by the factor CORPUS: the OBC and the CLMET behave significantly differently. As expected, another relevant factor is time, i.e. PERIOD. This is summarised in the regression output in Table 39, where the estimates represent the likelihood of *you were*. POLARITY, which was also coded for, was not a significant predictor.

	estimate *b*	**SE**	**z value**	**p-value**	**confidence intervals**	
					2.5%	**97.5%**
Intercept	0.37348	0.16826	2.220	<0.05	0.04882615	0.7096955
PERIOD= 1780-1850	1.50287	0.09352	16.071	<0.001	1.32116777	1.6878778
PERIOD= 1850-1913	4.86274	0.19808	24.549	<0.001	4.49068522	5.2695489
CORPUS= OBC	-1.49512	-8.547	-0.17493	<0.001	-1.84540801	-1.1587567
Concordance Index *C*		0.83				

Table 39. Output of logistic regression including predictors PERIOD and CORPUS; based on OBC and CLMET-drama

Figure 55 presents a visual representation of the effect of the factor CORPUS: globally, *you were* is preferred in both corpora, but the model predicts a higher probability of *you were* in the CLMET-drama (94.2% as opposed to 78.6% in the OBC).

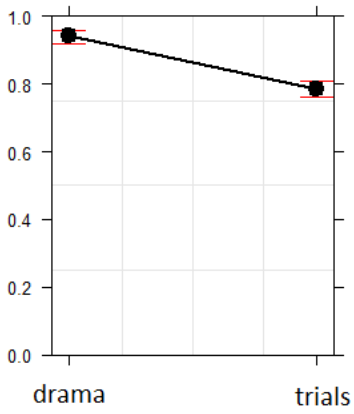


Figure 55. Effect of factor CORPUS on likelihood of *you were* (data: OBC and CLMET-drama)

In turn, that means that the chance of encountering *you was* is higher in the OBC.

A diachronic comparison of the two corpora in Figure 56 shows the same overall trend, i.e. *you was* increasingly being ousted by *you were*. However, it also reveals that the variant *you was* was never as widespread in the dramatic texts as in the trial proceedings.

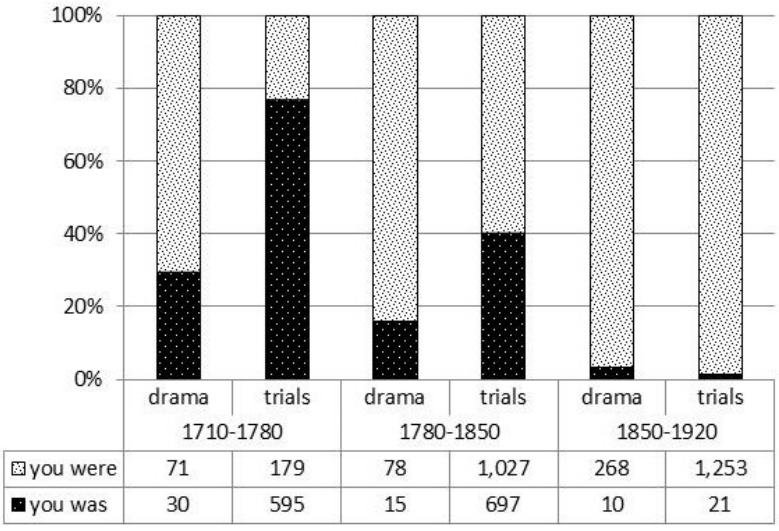


Figure 56. Proportions of *you was* and *you were* in the OBC and the CLMET-drama, by period (N = 4,244)

In the first two periods, the observed proportion of *you was* in the CLMET is markedly smaller than in the OBC (1710-1780: 29.7% compared to 76.9%; 1780-1850: 40.4% compared to 16.1%). In the final period, *you were* has been established as the clearly dominant variant in both corpora, with *you was* being used less the 5% of the time.

7.3.3 Conclusions

The analysis of *you was* and *you were* in the Late Modern period shows that *you was* is increasingly replaced by *you were*. Different genres show drastically different distributions.

The trial proceedings start with relatively high percentages of *you was* in the 18th century: it is the majority option. In the 19th century, *you were* predominates and *you was* drops below 10%. The regression line for the OBC data approximates an S-curve. In the dramatic texts in the CLMET, *you was* never reaches percentages as high as in the OBC, but the general downward trend is present as well. As *you was* only became a widespread option in the 18th century, it is not likely that the change from *you was* to *you were* as the favourite option was simply quicker in the dramatic texts. It would also be at odds with all the other developments so far discussed, where the trial proceedings always were ahead of the curve. Rather, it seems likely that *you was* never got as much of a foothold in dramatic texts as it did in trial proceedings. This may be an effect of the nature of these genres' relation to spoken language, which is assumed to be the point of origin for the form *you was*. Trial proceedings (ideally) record speakers' actual usage, while dramatic texts are invented and based on what the authors assume spoken language to be like. It is possible that authors in this case underestimated and underrepresented how widespread and frequent *you was* was. At the end of the Late Modern period, though, the CLMET actually retains a marginally higher rate of *you was* than the OBC. This pattern is familiar from the discussion of *I says / I said* in Chapter 6: in the 18th century, the OBC is more permeable to non-standard variants but in the final period (1850-1920), it is less so as an effect of increasing (linguistic) formality in the courtroom. In contrast to *I says*, there are no indications that *you was* was in any way affected by scribal interference.

As mentioned above, social factors could not be shown to have a significant impact on the distribution of the variants *you was* and *you were*. Instead, the variant comes to be less used by all groups of speakers in the course of time. It is plausible that *you was* simply was no longer acceptable in formal speech. The decline in other written speech-related genres like letters and drama also attests to this. While the growing standard ideology can be said to have an effect in this way, the impact of

grammatical prescription on more informal spoken language was probably not significant, as *you was* is still very frequent in present-day English.

7.4 Summary

After charting the development of the variants *you was* and *you were* in the OBC and the CLMET-drama, it is now possible to check findings against the hypotheses developed for this feature in 3.6.

The expected stable variation between *you was* and *you were* based on the present-day distribution of these variants in conversation, is not confirmed by the data. Instead, we find a clear changing preference from y*ou was* to *you were*. In the OBC, *you was* is very popular in the 18th century before its proportion drops in the 19th century. In the dramatic texts, *you was* is never that widespread but still exhibits a decline throughout LModE. These downward trends for *you was* are related to the fact that drama and trial proceedings are both representations of spoken language that underrepresent actual usage of *you was* for varying reasons, especially in the 19th century. In the OBC, the underrepresentation is likely due to the increasingly formal register in court, and the CLMET-drama probably contains less *you was* than spoken language because invented dialogue does not accurately represent the feature as it occurs in spoken conversation, where it is frequent until the present day.

The present analysis can confirm the expectation that developments will differ between genres/corpora: while the general trend is the same (from *you was* to *you were*), the OBC allows more room for variation in the early proceedings and less than the dramatic texts in the late proceedings. While it is true that the CLMET never matches the highest percentages of *you was* in the OBC, the lowest ratio of *you was* in the dramatic texts is still higher than the lowest OBC percentage. Taken together with the results on *I says* (Chapter 6), this points towards the OBC being more accepting of informal features in the early Late Modern period, but less so towards the end of the period.

The prediction that heavily stigmatised variants will be rare in trial transcripts due to the formality of the situation and the increasingly formal genre conventions is partly confirmed. In the 18th century, this tendency is not observable, but from the

1800s onwards, it is. At this point, *you were* was clearly established as a standard, high-prestige variant, while *you was* became a stigmatised option, which speakers apparently no longer felt comfortable using in this context.

8 Theoretical and methodological implications

> HENRY GURRIN. I have examined the various documents in this case [...]. (OBC, t19020909-644)

This concluding chapter will synthesise what can be learnt about language variation and change in the Late Modern period based on the preceding four case studies, and provide an answer to the three overarching research questions that were initially formulated (see 1.2 and 3.6):

A. How do variation and change manifest themselves in selected morphosyntactic features in Late Modern English with regard to the timing of change (if present) and its social and linguistic factors? How do different speech-related genres compare?
B. How are the variants evaluated in grammars of the time (positive / negative / changing)? Is there a correlation between this evaluation and actual use?
C. How suitable are the *Proceedings of the Old Bailey* (and trial proceedings in general) for historical sociolinguistics? How close to the dialogue uttered in the courtroom can we assume these published transcripts to be? What needs to be taken into account (e.g. in terms of scribal/editorial interference) when basing linguistic analyses on trial proceedings?

The present chapter provides insights into the social dimension of variation and change in answer to questions A and B (8.1), and discusses the issue of investigating speech via written sources in answer to question C (8.2). As these two sections outline all substantial outcomes of the present work, Section 8.3 will be restricted to brief concluding remarks, including opportunities for future work.

8.1 The social dimension of variation and change

This section summarizes key findings on the effects of social factors on language variation and change in Late Modern morphosyntax. The social dimension of variation and change is understood to include both the impact of social factors on language choices and the contemporary social evaluation of variants. After a review of the

findings in 8.1.1, their theoretical and methodological implications are explored in 8.1.2.

8.1.1 Review of findings

Out of the four features under investigation, those two involved in change – the change from BE to HAVE as the perfect auxiliary (*he **is** just come home – he **has** just come home*) and from MUST to HAVE TO as the preferred obligation marker (*you **must** go now – you **have to** go now*) – show a significant impact of social factors. Both record a social class effect on their distributions, though the social group leading the change differs. Gender does not significantly influence the distribution of any variable.

Lower-class speakers led the change towards HAVE TO. Previous research has advanced the notion that HAVE TO originated in informal conversation, first spread in informal texts and ultimately was a change from below the level of consciousness. In the grammars of the time, discussion of variation between MUST and HAVE TO was certainly rare. The newer variant HAVE TO is hardly even acknowledged. On the one hand, this can be read as conservatism, but on the other hand, it means that no strong opposition to HAVE TO was present. This is further indication that the change was a change from below.

In the change from BE + participle to HAVE + participle, the higher social classes were ahead of the change. It is difficult to explain why this might have been. HAVE + participle is not known as a variant that originated in more formal, educated registers and thus might have been more widespread among the higher classes who had better access to education and were more exposed to formal registers. A direct influence of grammar writing can be excluded, in any case: grammarians in the 19th century widely acknowledge variability between HAVE and BE, and the first comments that actively advocate HAVE appear only in the 1820s. While it is true that BE + participle is strongly discouraged in subsequent decades and that we thus witness a much stronger evaluation than in the case of MUST and HAVE TO, these comments do not precede changing usage, but follow it. The data from the OBC show that all participles of mutative intransitives except *gone* already strongly favoured HAVE by then. The grammars thus largely chart already established usage.

In both cases, it has to be said that the influence of class is significant but rather small in comparison to the other factors that could be established as significant. Figure 57 and Figure 58 display dotplots based on so-called random forests,[159] which show variables ordered according to their importance. Random forests are advocated in Tagliamonte (2012: 153) as an ideal tool to expose "the relative contribution (i.e. strength) of each factor group on the variable under investigation".

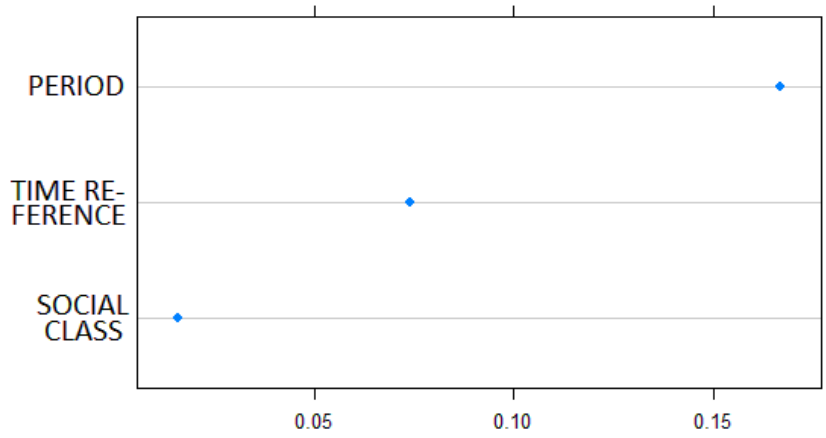


Figure 57. Strength of effects on variation between MUST and HAVE TO, OBC

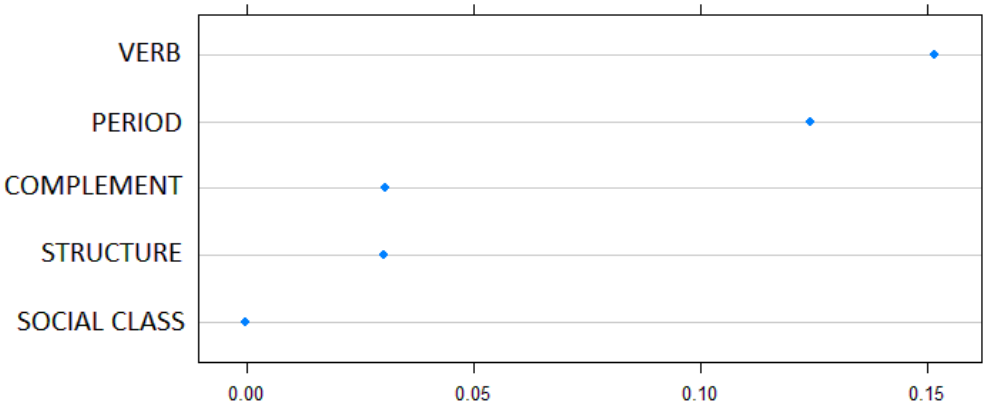


Figure 58. Strength of effects on variation between BE perfect and HAVE perfect, OBC

[159] A random forest is a non-parametric alternative to regression modelling. It yields the importance measure for every variable in a model averaged over many conditional inference trees. A random forest can be created using the *cforest()* function in the R package *party*, and the dotplots can be created using the function *dotplot()*. I used the procedure explained in Tagliamonte (2012: 152–153) to create these dotplots illustrating variable importance. It needs to be stated that conditional inference trees, which underlie random forests, are created with an algorithm based on permutation (drawing numerous random samples from the original sample), which is why the results are slightly different every time the code is executed (Levshina 2015: 297–298).

It is evident that SOCIAL CLASS is the least important factor for both variables. Time (here represented by the factor PERIOD) and linguistic factors are much more important. The social class effect, while clearly indicated by the logistic regressions run for these two variables, should therefore not be overstated.

For the other two variables, *I says / I said* and *you was / you were*, the social factors gender and class had no significant effect. Figure 59 depicts the effects of significant factors in the distribution of *I says / I said*.

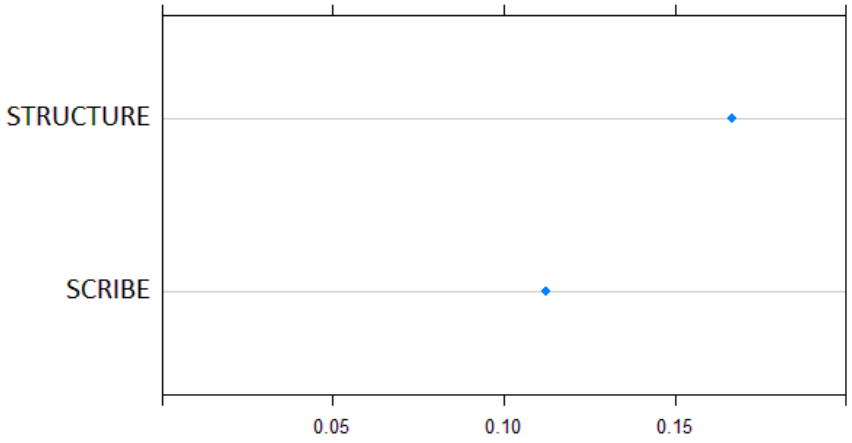


Figure 59. Strength of effects on variation between *I says* and *I said*, OBC

Figure 60 illustrates significant predictors for the variation between *you was* and *you were*.

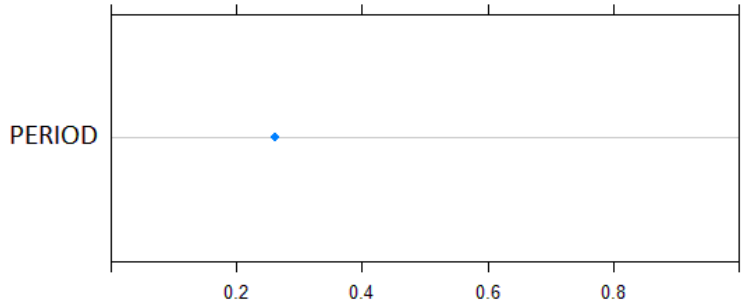


Figure 60. Strength of effects on variation between *you was* and *you were*, OBC

For variation between *I says* and *I said*, another significant social factor – although located above the level of individual speakers – emerges: SCRIBE (further discussion in 8.2). Even more of an impact is made by the linguistic factor STRUCTURE (i.e. whether the pronoun or the verb comes first), which turns out to be an important clue

to the reasons for the overall decline of *I says*. Except for the brief period in the mid-18th century in which scribal interference must be assumed, dropping numbers for *I says* are mainly interrelated with a different process. The disappearance of inverted verb-pronoun constructions (e.g. *says I, says she*) from the *Proceedings*, which were a context overwhelmingly associated with *says* (instead of *said*), play a role in this respect. For *was/were* variation (Figure 60), only time (represented by the factor PERIOD) could be shown to have a significant effect on the variable's development.

Interestingly, both the *was/were* and the *says/said* variable contain one strongly stigmatised variant. i.e. *I says* and *you was*, respectively. As concord was one of the most discussed topics in the grammars of the time, and any deviation from standard concord heavily stigmatised, a significant dispreference for these variants among higher-class educated speakers would have been expected. Instead, *you was* was the preferred variant across the board in the 18th century and almost completely disappears in all social groups in the 19th century. The situation is broadly the same for *I says* (except for the era of scribal interference). For both variants – which obviously are both still found in contemporary conversation – we find extensive prescriptive judgement, but no class difference and a decline that comes much later than the critical comments. It seems that changing ideas of what constituted formal language and changing standards of acceptability in courtroom discourse motivated the exclusion of *you was* and *I says* in 19th century courtroom dialogue. Speakers' register consciousness rather than some more global prescriptive effect appears to be at the heart of this development.

Prescription does not serve as a fitting explanation in the two change processes, either. Instead, we witness the following two scenarios: for MUST/HAVE TO, there is (practically) no prescriptive comment during the change, and for BE/HAVE + participle, prescription follows or is at least simultaneous with changing practice. Previous work on prescription in Late Modern English verbs and verb categories has also come to the conclusion that there is "only little evidence of an influence of prescriptive grammar writing on actual language change" (Anderwald 2016: 245). Therefore, I would not directly ascribe speakers' choices to grammatical prescription.

In contrast to social class, the factor gender was never flagged as significant in the analysis of the OBC material, although it could have been expected. For instance, it

would have consistent with earlier research that women are more conservative in changes towards an innovative variant that is initially criticised. The analyses in this study showed that linguistic factors in general had a stronger impact on variables than social factors. This begs the question whether social factors play as influential a role as initially assumed for these four morphosyntactic variables. The theoretical and methodological implications of the above findings will be explored in the next section.

8.1.2 Implications of the findings

It is by no means the case that we encounter familiar sociolinguistic patterns in this analysis or that we can establish generalisations about social processes across all four variables. This does not automatically mean that social factors do not play a role for variation and change processes associated with these variables, but it does require some further discussion. Several explanations present themselves: It is, of course, possible that we need to rethink established patterns and their applicability to the Late Modern period. It is also possible that morphosyntactic features are less likely to be imbued with social meaning than, say, phonological variables. Finally, the social patterns found could be a reflection of the analysed textual material rather than a reflection of linguistic reality. In the following, these three potential explanations will be briefly discussed.

Concerning any disparity between the present results and previously observed social patterns (e.g. middle classes leading in changes from below, women being more hesitant to employ stigmatised variants), it was already initially stated (see 2.3.1) that we should not depend too heavily on so-called established patterns, as we cannot generalise across communities and periods.[160] For the variables in question, there are comparatively few studies of Late Modern usage. If we focus on these, though, the present results fit quite well with the information gathered so far. In the case of MUST and HAVE TO, our results do not contradict the hypothesis that HAVE TO was established via change from below in the 19th and 20th centuries. The change was little commented on in grammars, for instance. For the change from BE + participle to HAVE + participle with mutative intransitives as well as for the variation between *I says* and *I said*, no

[160] In a study of Late Modern relative clauses, also based on the OBC, Huber (2017: 112) points out that some of his findings "question earlier assumptions about the development of English RCs [relative clauses] and about the social mechanisms of language change in general".

comparable Late Modern studies of social factors are available. The only variable that fails to match a previously established social pattern for Late Modern England is *you was / you were*: the gender effect reported in Laitinen (2009), i.e. that *you was* was initially propelled by men, but also first abandoned by men, is not confirmed by the OBC results. However, Laitinen (2009) bases these conclusions on rather small absolute numbers, thereby possibly overstating the impact of gender (see 7.3.1).

The second explanation attempt, i.e. that morphosyntactic variables are less prone to be sociolinguistically conditioned, has been put forward by a number of scholars: In a discussion of the state of research on men's and women's language choices and the scarcity of work addressing syntactic variables and gender, Cheshire (2002: 439) suggests that syntactic constructions like BE / HAVE + participle are "unlikely to occur frequently enough to become habitually associated with the speech of either women or men". However, there are arguments to the contrary, too: Tagliamonte & Baayen (2012: 171), for instance, argue that the morphological contrasts which occur as part of grammatical change, "appear to be especially amenable to the embedding of social meaning". *Was/were* variation is even cited as one such example. Chambers (2009: 56–57) asserts that grammatical variables function as widespread class markers in present-day English (also see Nevalainen et al. 2011: 3). That they would also do so in Late Modern English, an era marked by codification and growing social evaluation of language, is a reasonable assumption.

A final explanation for the results of the present study focuses on textual and methodological factors rather than calling into question the importance of social factors in general. After all, the impact of social factors has been continually shown in earlier work – also in work based on other syntactic features in the OBC, such as relativizers (see Huber 2017). In particular, the following aspects deserve consideration: the nature of the corpora and the individual texts that were used, the choice of social categories for analysis and the methods of analysis. While the texts under investigation, especially the *Proceedings*, have already been discussed extensively in 3.2, some related issues that emerged in the course of the analysis deserve a closer look in this concluding chapter – also in light of possible future work.

That women and the lower classes are underrepresented in the material was already apparent from the start, but in some extreme cases the effects of this

circumstance were more problematic than predicted: when analysing *you was/were*, represented by thousands of tokens in the large OBC, so few data points were provided by women that further (statistical) analysis was not effectively possible and conclusions on the impact of social factors had to remain limited (see discussion in 7.3.1). This essentially means that an effect of social factors might very well be there in reality, but cannot be reliably verified. This problem is not to be underestimated, especially as research on present-day English has suggested that linguistic gender differentiation is greatest among those parts of the population in which power is scarcest and "where women's access to power is the greatest threat to men", i.e. in the lower classes (Eckert 1989: 256). It is just this segment of society for which data is most difficult to come by in the Late Modern period. While the OBC allows for a very good – sometimes even unprecedented – opportunity to investigate Late Modern morphoyntax under social considerations, any analysis will be impacted by this issue. In addition, scribal interference (dealt with in detail in 8.2) made analysis more than difficult for some variables, and also prevents us from making stronger claims on social variation.

Another issue concerning the nature of texts is the fact that speakers in court proceedings very often report what other people said, and thus to some degree make use of the words of others rather than their own: in example (75), DS John Mulvaney reports the speech of DI Thompson, his superior, and also that of a suspect in a forgery investigation, Mr. Holchester. He even seems to try and recall the exact words used (see the wording "notes" vs. "Russian notes").

(75) JOHN MULVANEY (Detective Police-sergeant). [...] I saw some coloured paper in the grate. [...] I took it out, opened it, and there were five notes of five roubles each of the Bank of Russia. [...] [Inspector Thompson] put them on the table [...] and said, "You see here is a good bundle of notes;" and [the alleged forger] Holchester said, "I don't know Russian; I never saw a Russian note in my life". I don't remember whether Thompson said "Russian notes;" he said, "Here is a good bundle of notes." (OBC, t-18651218-140)

Such texts leave corpus compilers in the difficult position of having to decide how to treat 'speech-in-speech' utterances like those by Mr. Thompson and Mr. Holchester. Should speech within another speech event be tagged as a separate speech event and

annotated with social information on the cited original speaker, as it (presumably) represents their words? Or should the entire speech event be ascribed to the speaker who reports speech by another person (i.e. DS Mulvaney in the present example)?

In the OBC, the second option, i.e. annotating only 'top-level' speech with the respective sociobiographical information, was used.[161] This makes good sense, too: first of all, it is very difficult to decide where to draw the line when identifying speech events within other speech events: Should only direct speech be considered? Should other forms like free direct or free indirect speech also be taken as separate speech events? Secondly, there is obviously a limit to the amount of time, effort and financial means that can be devoted to such aspects in corpus compilation. Considering the size of the OBC and the fact that sociolinguistic tagging was very labour-intensive, it was decided not to annotate speech events within speech events. However, this solution certainly omits some information about who (supposedly) said what or used a certain linguistic construction. Theoretically, results may be impacted by this.[162]

Despite ample evidence in earlier work that social components play a role, only a limited impact of the social factors gender and class can be shown in the present investigation. It bears repeating that this does not mean that social variation was nonexistent. Instead, it may have manifested along different lines, which were not accessible via the chosen method of analysis. For instance, aspects could play a role that could not be investigated in detail: Priming effects (see 3.5), for instance, could have impacted the choice of variants in the dialogic exchanges in court. It is further possible that a person's role in the courtroom or the addressee of an utterance were important with respect to variant choice.

As far as priming is concerned, it was not feasible to efficiently integrate this factor into the analysis of several multi-million word corpora in the present study. It was therefore not considered. The factor role, while available in the OBC, could only be used for analysis in a limited manner. The data on speaker roles in the OBC 1.0, come with some built-in restrictions due to the historical context of the *Proceedings*: especially collinearity issues caused by the pairing of professional roles (lawyer, judge)

[161] This information is known to me because I was part of the team compiling the OBC 1.0.

[162] This is arguably a bigger issue for other types of research questions: Widlitzki & Huber (2016) actually include a discussion of 'top-level' annotation vs. speech-in-speech annotation for sociobiographical factors. The study is concerned with swearing and taboo language in the *Proceedings*, which practically only occurs in the courtroom in witness testimony on what other people said. A small subset of the tokens was recoded by hand so that the information on the 'original speakers' was available for the analysis.

exclusively with higher-class male speakers is a problem. As for the impact of addressees, previous research has indicated that they play an important role for linguistic gender differences (see e.g. Biber et al. 1998: 216). It was not possible to integrate the variable addressee into the analysis based on the present corpus, but perhaps this may be an avenue worth exploring in the future. Finally, it also bears repeating that the social dimension of variation and change is represented by more than a link between 'traditional' social factors like class with linguistic variant choice, but can for instance include aspects like migration movements or social networks. Future work will surely complement the present results by including such aspects, and in this way add to our understanding.

8.2 Tracing speech in historical writing

This section summarizes key issues in investigating speech via written historical texts. Section 8.2.1 discusses the development of the four variables in different genres and what this tells us about different kinds of writing. In Section 8.2.2, theoretical and methodological implications for historical sociolinguistics are discussed.

8.2.1 Review of findings

A subsidiary aim of this study was to assess the potential of the OBC for historical sociolinguistic research (research question C). To establish this, it was necessary to ask the question "what are written texts representing spoken face-to-face interaction like?" (Culpeper & Kytö 2010: 3), and more concretely, to ask what the P*roceedings of the Old Bailey* are like. This section will review findings on 'external fit' and 'internal consistency' of the OBC, i.e. how the corpus compares to other Late Modern corpora and how the *Proceedings* developed as a genre.

In order to assess the OBC's external fit, the results from the trial proceedings were compared to results from another speech-related corpus - in three cases, to the drama subsection of the CLMET, and in one case to the narrative fiction subsection of the CLMET. For all four variables, significant differences between the OBC and the CLMET were found. This confirms once again that genre is a major factor in language variation and change. Figure 61 and Figure 62 show overviews of the analysed change

processes (from MUST to HAVE TO, and from BE + past participle to HAVE + past participle).

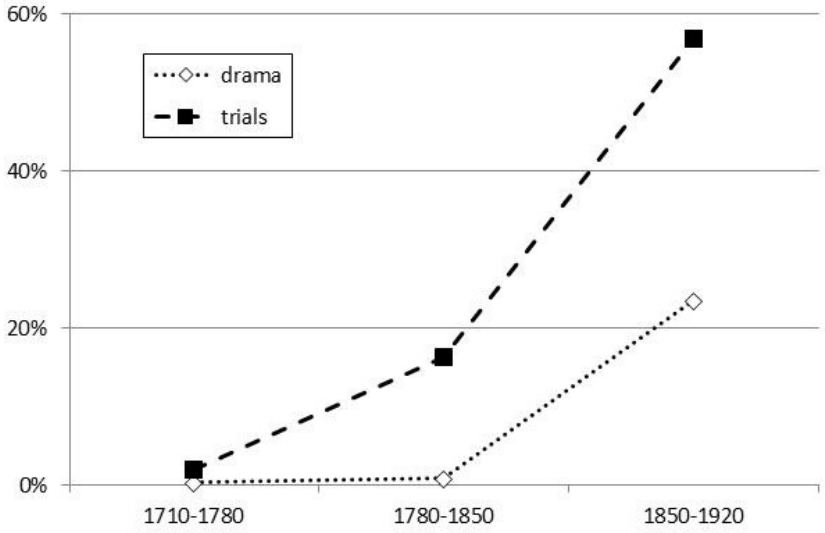


Figure 61. The change from MUST to HAVE TO: percentage of HAVE TO in the OBC and the CLMET-drama

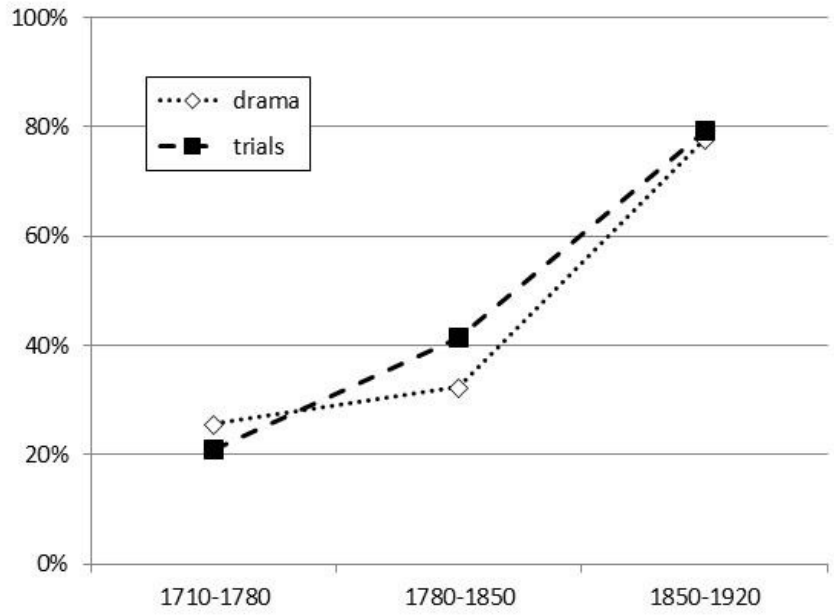


Figure 62. The change from BE + PP to HAVE + PP: percentage of HAVE + PP in the OBC and the CLMET-drama

In both cases, the development in the trials is ahead of the drama corpus, at least in periods 2 and 3. Granted, this advantage is rather slight in the case of BE/HAVE +

participles of MIs. This process is closer to completion than the change from MUST to HAVE TO, which is still ongoing in contemporary English. In the case of the newer change (MUST > HAVE TO), I would argue that the transcribed speech in the OBC is farther ahead than the dramatic texts because it more directly reflects changes in spoken trends for this variable than the dramatic texts, which rely on authors' intuitions on what is widespread in a speech community. In the case of BE and HAVE, the dramatic genre already had some time to catch up to the development in spoken usage. Figures therefore more closely match those in the trial proceedings, which more closely mirror spoken usage for these variables.

The two variables *I says/said* and *you was/were* were not involved in global language change in the Late Modern period, but, as it turned out, nevertheless showed a less than stable distribution in the corpora under investigation. Their developments in the OBC and the CLMET are shown in Figure 63 and Figure 64.

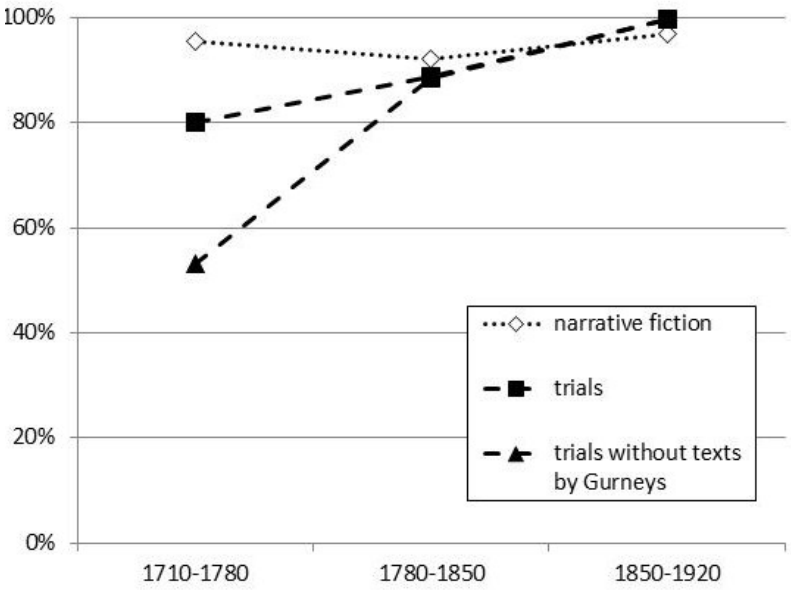


Figure 63. Variation between *says* and *said*: percentage of *said* in the OBC and the CLMET-narrfic

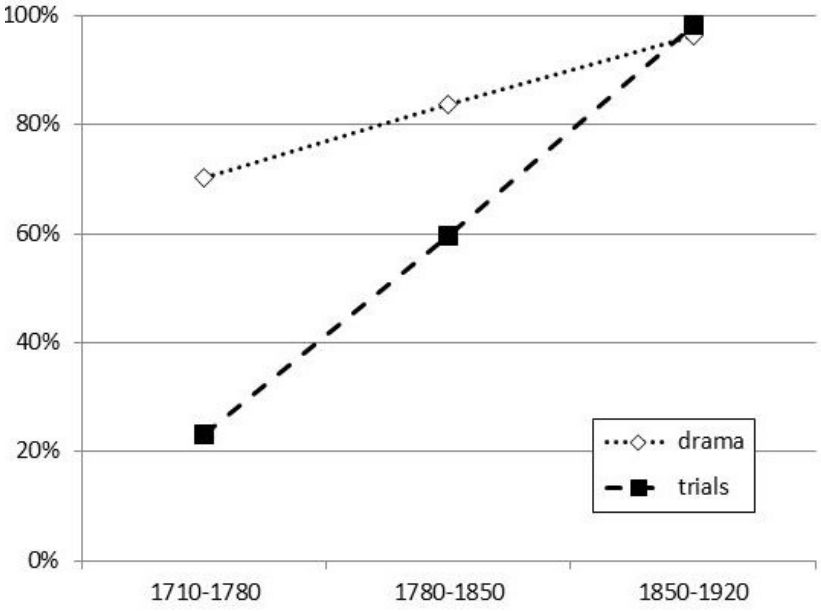


Figure 64. Variation between *you was* and *you were*: percentage of *you were* in the OBC and the CLMET-drama

As the figures indicate, the OBC's development relative to the respective CLMET-subcorpus is similar for both variables: originally showing a comparatively high proportion of the more informal variants, the OBC subsequently shows a steep rise of the more formal 'standard' variants (*were* and *said* respectively). For *says/said*, scribal interference by the Gurney family probably artificially inflated the proportion of the variant *said* in the first two periods under observation. For this reason, the diagram in Figure 63 contains an additional line (called 'trials without texts by Gurneys'), which indicates the hypothetical developments without the assumed scribal interference: according to this adjusted trajectory, *I said* was less frequent in the beginning, but a steep rise took place nevertheless.

This pattern of initially high incidence of the non-standard concord variant followed by a steep decline is not found to the same extent in the CLMET. For *says/said*, the development in the CLMET shows a rather constant proportion of *said* at a very high level (95.6%, 92.1% and 96.9%). The CLMET thus shows the stable proportions through time that would be expected of longstanding variation. After all, both variants are well attested in spoken conversations today. However, the proportion of *said* is higher than I expected and much higher than it is in the OBC for most of the period under analysis. Informal *says* largely had better chances of being heard in the

courtroom than being found in dramatic dialogue. That the OBC does not show stable variation is an indication that the trial genre is undergoing change towards the exclusion of informal variants. Concerning *was/were*, the CLMET does not show the expected stable variation between the two variants but a trend towards *were*, which brings this development closer in line with results from the OBC. However, *were* was already much more strongly entrenched as the favoured variant in the CLMET from the beginning (more than 70% *were* compared to ca. 23% in the trials). In the final period, both corpora show close to 100% *were* usage (96.4% in trials, 98.4% in drama). The variant *you was* constituted a broadly accepted, even fashionable option for only a limited period of time. Apparently, it did not retain that status for long enough to be widely used in all types of speech-related texts. While the trial proceedings show a large proportion of *you was* in the 18th century, *you was* never gained a strong foothold in dramatic dialogue.

These comparisons of the results for *you was/were* and *I says/said* across the two corpora give rise to the supposition that the OBC is closer to spoken informal norms in the beginning of the Late Modern period, as e.g. shown by the high tolerance of *you was* and *I says*, but over-represents standardising tendencies later on, especially in the late 19th century, as e.g. shown by these variants dropping to close to zero. This means that the OBC does not display internal consistency for the variables investigated here. What is at the root of this? Scribal interference can only explain a small portion of the phenomenon: except for the mid-1700s, when the Gurneys actively suppressed *I says*, there is no compelling evidence that either scribes, printers or editors globally and systematically made 'corrections'. The simple fact that variation is observed for both *was/were* and *says/said* from the beginning to the end of the *Proceedings* is evidence against such a hypothesis. It is further interesting that the change from *says* to *said*, which I assume some scribes made, is a very small orthographic change. It makes sense that scribal interference would only take place for such small alterations, like 'fixing' the verb ending.[163]

What seems to be far more important than corrections or interference at the level of the written publication are the changing courtroom procedure and the changing

[163] Findings for other variables support this idea: Huber (2010: 353) argues that the declining rate of negative contraction in the *Old Bailey Proceedings* was partly due to external pressure by the City of London, which increased its control of the publication throughout the 18th century: as a result, the formality of the texts increased, at the same time reducing the rate of variants characteristic of spoken language.

register in court. For the most part, I assume that the move towards the exclusion of informal, conversational variants was due to speakers using them less frequently in the courtroom setting as the whole process was becoming more formalised, the role of lawyers became more and more important and witnesses were increasingly better prepared before appearing in court. Linguistic flexibility decreased in other ways, too: a major development in this regard was the fixing of conversational roles. In Early Modern trials, conversational roles like 'questioner' or 'initiator' were still flexibly assigned, as e.g. Archer (2005) notes: defendants and witnesses, for instance, could also be 'initiators' and 'questioners', not only judges, who more typically assume these roles. However, these roles become "non-transferable marker[s] of power"[164] in Late Modern trials (Archer 2005: 287).

8.2.2 Implications of the findings

This study has shown differences between genres and differences within the genre of trial proceedings throughout time. It has also provided evidence for scribal interference in the *Proceedings*. To discuss these findings with regard to their theoretical and methodological implications, the present section addresses the relationship between courtroom dialogue and published transcripts, the development of genres, and finally the usefulness of the *Proceedings of the Old Bailey* and trial proceedings in general for historical sociolinguistics.

Examining the relationship that holds between historical spoken interaction and the record we have of it has been part of historical sociolinguistics since its inception. This is primarily relevant when a study aims to shed light on speech in past stages of English. Although this study does not exclusively focus on this aspect, the connection between courtroom dialogue and court transcript is relevant for the interpretation of the results concerning informal or conversational features (like *you was* and *I says*). While only the case of *I says* shows evidence of deliberate interference, it is clear that the possibility of distortion – whether deliberate or an effect of the transition from one mode to another – is present at any point after the spoken word is uttered. The spoken material recorded in the *Proceedings* undergoes a number of stages (shorthand transcription, manuscript, typesetting, perhaps proofreading, printing) and the written

[164] Archer (2005) points out that the term "non-transferable marker[s] of power" originates in Walker (1987: 62).

text is thus quite removed from the spoken word, both temporally and in terms of the number of intermediate stages involved. No easy conclusions on 'quality' for linguistic analysis can be drawn from just that, though. Rather, it seems that sometimes the process of publication may have hindered the accurate recording of spoken material (e.g. scribal interference with *I says*), but sometimes the process may have been advantageous (regulations to report accurately may have occasioned the reappearance of *I says* in the proceedings).[165]

It is important to note that the texts in the OBC (and the larger genre of trial proceedings) change through time, and that their relationship to the underlying original speech events is also subject to diachronic change. The results on informal, conversational features suggest that the OBC moved away from conversational to more formalised speech in the course of the Late Modern period because the register conventions in the courtroom changed in this direction. For the features presented here, I argue that register changes rather than genre changes were responsible for the more formal language in later periods. With the formalisation and professionalisation of trial procedure, individuals in court make use of increasingly formal language. I would only attribute a greater impact to another factor in the case of *I says*, which was suppressed temporarily in the 18th century by scribes. This is attributable not to changes in people's speech but to scribes' understanding of what was appropriate for a written record of a trial – and thus happens on the level of genre. The OBC is – like all written sources – of course less oral in its characterisation than speech. Additionally, it seems to underrepresent conversational features especially in the latter periods – and thus, to less quickly take over changes originating in the spoken language.

It is useful at this point to briefly return to Krug's visualisation of innovation diffusion in changes from below as several S-curves representing different genres, shown in Figure 3 and repeated here as Figure 65 for convenience.

[165] Studies on contractions in the OBC (Huber 2010: 69) and taboo language in the OBC (Widlitzki & Huber 2016: 330) show a resurgence of contracted forms and 'bad' language, respectively, in the late 19th century after periods of decline from about the 1750s onwards. One argument that is put forward is that the "portrayal of the spoken word became more faithful again in the second half of the nineteenth century" (Widlitzki & Huber 2016: 330).

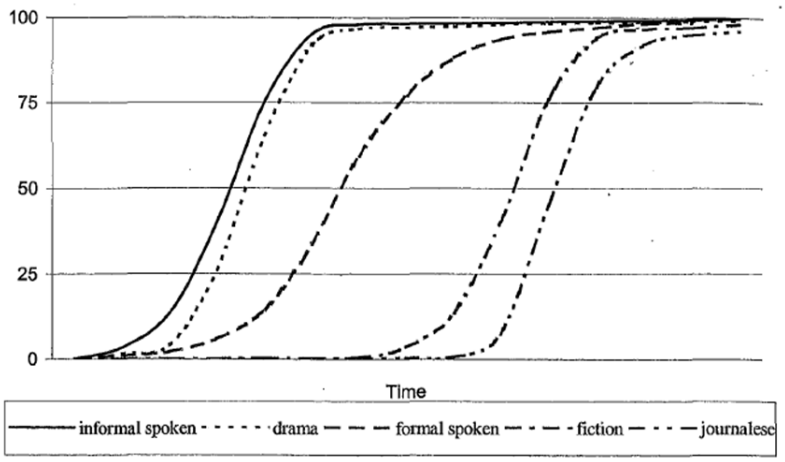


Figure 65. Innovation diffusion in spoken and written language, after Krug (2000: 196)

Of course, reality is much more complex – which Krug (2000: 198) doesn't dispute: he acknowledges that there may be intersections between curves as genres develop towards either the more spoken or more written end of the continuum he postulates. The present analysis plainly underscores the necessity to integrate the possibility that a curve associated with a particular genre may move further to the right or left in time. Trial proceedings (which would be part of 'formal spoken' in Krug's diagram) were shown to be much closer to the spoken end of the continuum in the beginning of the Late Modern period than dramatic texts. At this time, they readily took up innovations from the spoken language. However, this 'headstart' in comparison to dramatic or narrative texts diminished later on. Towards the end of the Late Modern period, the situation is closer to what Figure 65 depicts: trials are now more formal, and dramatic texts are more open to innovations from speech.

The model runs into trouble in another respect: individual linguistic features (such as the variable HAVE TO/MUST or the variable *I says / I said*) or bundles of features may behave independently of each other within the same genre. Obviously, it would diminish the explanatory value of a model to include a curve for each feature. This is why I find Leech's (2013: 114) idea of prestige barriers useful: prestige barriers impede or allow the progress of a linguistic feature into certain texts. I would like to add to this idea and argue that prestige barriers can become stronger or weaker throughout time, and that prestige barriers exist not only for written texts but also for spoken registers. It is further important to note that the prestige barriers that hold for a

certain genre (such as court proceedings) and a register associated with that genre (such as courtroom discourse) may differ. In practice, though, and especially with less eye-catching features (like the grammatical variables discussed here), disentangling the dynamics of genre from those of register can of course be very difficult if not impossible.

In general, the trial proceedings as a whole are at least as amenable to variation as other speech-related text types used in this study and in earlier work. Like the 19th-century trials investigated in Kytö & Smitterberg (2006: 209), the *Proceedings* display "features from present-day conversation but were also characterised by language use typical of the courtroom situation". They not only provide valuable information on the genre of trial proceedings, but also are valuable sources for the investigation of so-called conversational / informal features. For instance, the analysis of *you was* has shown that it was much more frequent in the OBC than in other speech-related texts, and *I says*, considered "almost entirely restricted to conversation" (Rühlemann 2007: 170), could be retrieved 1,150 times from the OBC. The OBC offers possibilities to extract and investigate phenomena like these, which are considered very 'spoken' or rare.

Of course, trial proceedings are more useful for some variables than for others, simply due to their nature. This issue has been extensively addressed in the sections on methodological background in Chapter 3 and in the course of the analysis. To name just one example, let us briefly return to the analysis of the variable *I says/said*. To retrieve tokens, past narrative is needed. Trials should therefore be ideal for obtaining this kind of data. It is important to note, though, that not all trial participants provide past narrative and that the variable will therefore be practically absent in the utterances of lawyers and judges. This applies to all corpora including trial transcripts from that era. It is important to keep this issue in mind when reporting results, especially when normalising: a relative frequency with the entire OBC as a basis would be less useful in this case than a normalised frequency with only speech by those participants who provide past narrative. Other variables are distributed differently: questions, for instance, are almost entirely restricted to the speech of judges and lawyers.

Finally, selecting texts in historical sociolinguistics is of course also informed by practical aspects, not just considerations of their quality, accuracy, potential to

inform about spoken usage, inclusion of variables the study focuses on, etc. On a much more basic level, the texts need to be available in sufficient quantities to allow for analysis, and social information on writers or speakers must be available (Culpeper & Kytö 2010: 15). The OBC meets these requirements and provides valuable data for historical sociolinguistics. If studies based on the OBC or similar corpora are undertaken with the necessary acknowledgment of their limitations, these corpora of trial transcripts can truly address the hope expressed in Tieken-Boon van Ostade (2000: 446) that Late Modern trial proceedings can "throw interesting light on how people really spoke" and "help us to get a fuller picture of the stylistic range available to speakers from the period".[166]

8.3 Concluding remarks and outlook

The present work has shown the development of four morphosyntactic variables in Late Modern English. It represents the first large-scale sociohistorical investigation into these variables in the 18th and 19th century, providing a necessary extension of earlier work on a smaller scale (as represented e.g. by Laitinen 2009 on *you was/were*) or as a complement to more long-term studies including more variables or variants (such as Krug 2000 on HAVE TO, MUST and other (semi-)modals). The results show that there are certainly interesting developments in the grammatical system of the Late Modern period, proving once again that the longstanding myth of stasis in this period is not tenable. It emerged that grammatical variables can be socially conditioned and are subject to social evaluation. Further insights include that prescription per se cannot provide an explanation for the developments observed, and that the impact of genre and register plays a great role. Methodologically, integrating the historical context and finding individual solutions for the analytical challenges provided by individual features are key to a successful analysis.

Beyond addressing the research questions formulated in the beginning, the analyses have also turned up some fascinating related results. For instance, the investigation of MUST and HAVE TO furnishes evidence that HAVE TO is favoured in contexts with past-time reference, and that the proportion of HAVE TO was consistently

[166] Tieken-Boon van Ostade actually refers to the 18th century in particular in this quote, but it makes sense to extend her statement to the 19th century.

larger in past than in present contexts throughout the change process. The dubious status of MUST in such contexts (still acceptable, especially in backshifting contexts, but becoming rarer) apparently propelled the rise of the alterative HAVE TO. For *I says*, it was established that its disappearance was associated with the disappearance of the inverted construction SAY + pronoun. These comparatively small insights play an important role in our efforts to understand the overall development of individual linguistic features (like the four morphosyntactic features discussed in the present study). Like pieces in a mosaic, they add to the larger overall picture. In combination with other individual efforts, they ultimately contribute to our understanding of variation and change on a larger scale. It is my sincere hope that the present investigation can assist others working on their own pieces of the linguistic picture.

Based on the present study, some opportunities for future work that could further the exploration Late Modern English sociolinguistics, come to my mind. Expanding on the ideas outlined in 8.1.2, it would be useful to create a 'Late Modern sociopragmatic corpus', perhaps similar in set-up to the Early Modern 'Sociopragmatic Corpus' (Culpeper & Archer 2007), described in Archer (2005: 107). It could, among other things, allow the integration of the addressee of an utterance and their social characteristics into the analysis and may help uncover hitherto hidden social patterns. The basis for such a corpus could be a small balanced subset of the OBC texts, with men and women of different classes equally represented. Such a smaller corpus would also more easily allow for the integration of factors like priming effects. On a small scale, it could also be rewarding to tag speech within speech (see discussion in 8.1.2). These measures would open up new opportunities to research the interaction and impact of social factors.

While it is true that most speakers in the history of English "have left not a single trace to document the words they spoke, or the conversations in which they participated" (Mugglestone 2006: 2), analysing transcribed speech and other speech-related texts gives us a much better idea of what their words and conversations could have been like. And researching what was involved in turning these words into the written documents available to us helps us understand their world a little better. When Milroy (2012: 583) says that "[t]he true history of a language is necessarily a social history, and therefore sociolinguistic insights can contribute enormously to it", this is a

definite call for further studies which put the speakers in their historical setting at the heart of the investigation of language. A lot remains to be done in historical sociolinguistics.

References

Aarts, Bas, Joanne Close and Sean Wallis (2013). "Choices over time: methodological issues in investigating current change." Eds. Bas Aarts, Joanne Close, Geoffrey Leech and Sean Wallis, *The verb phrase in English*: *Investigating recent language change with corpora*. Cambridge: Cambridge University Press. 14–45.

Aarts, Bas, María J. López-Couso and Belén Méndez-Naya (2012). "Late Modern English: Syntax." Eds. Alexander Bergs and Laurel J. Brinton, *English historical linguistics*: *An international handbook*. Berlin: De Gruyter Mouton. 869–887.

Aarts, Bas, Sean Wallis and Jill Bowie (2014). "Profiling the English verb phrase over time: modal patterns." Eds. Irma Taavitsainen, Merja Kytö, Claudia Claridge and Jeremy Smith, *Developments in English*: *Expanding electronic evidence*. Cambridge: Cambridge University Press. 48–76.

Addison, Joseph (1712). "Adde tot egregias urbes, operumque laborem…" *Spectator* (415).

Ainsworth, Claire (2015). "Sex redefined." *Nature* 518 (7539): 288–291.

Allen, Alexander and James Cornwell (1841). *A New English Grammar, with Very Copious Exercises, and a Systematic View of the Formation and Derivation of Words.* London: Simpkin, Marshall, & Co.

Allen, William (1824). *An English Grammar; with Exercises, Notes, and Questions* [1813], 3rd edn. London: G. and W. B. Whittaker.

Altick, Richard D. (1998). *The English common reader*: *A social history of the mass reading public, 1800-1900* [1957], 2nd edn. Columbus: Ohio State University Press.

Ammon, Ulrich, Norbert Dittmar and Klaus J. Mattheier, eds. (1988). *Sociolinguistics: An International Handbook of the Science of Language and Society*. Berlin: de Gruyter.

Anderwald, Lieselotte (n.d.). *Collection of Nineteenth-Century Grammars (CNG).* <http://www.anglistik.uni-kiel.de/de/fachgebiete/linguistik/anderwald/cng-collection-of-nineteenth-century-grammar>.

Anderwald, Lieselotte (2001). "*Was/Were*-variation in non-standard British English today." *English World-Wide* 22 (1): 1–21.

Anderwald, Lieselotte (2012). "Clumsy, awkward or having a peculiar propriety? Prescriptive judgements and language change in the 19th century." *Language Sciences* 34 (1): 28–53.

Anderwald, Lieselotte (2014a). "Measuring the success of prescriptivism: quantitative grammaticography, corpus linguistics and the progressive passive." *English Language and Linguistics* 18 (1): 1–21.

Anderwald, Lieselotte (2014b). "The decline of the BE-perfect, linguistic relativity, and grammar writing in the nineteenth century." Ed. Marianne Hundt, *Late Modern English syntax*. Cambridge: Cambridge University Press. 13–37.

Anderwald, Lieselotte (2016). *Language between description and prescription*: *Verbs and verb categories in nineteenth-century grammars of English.* Oxford: Oxford University Press.

Anderwald, Lieselotte (2017). "'Vernacular Universals' in nineteenth-century grammar writing." Eds. Tanja Säily, Arja Nurmi, Minna Palander-Collin and Anita Auer, *Future Paths for Historical Sociolinguistics: Methods, Materials, Theory*. Amsterdam: Benjamins.

Anthony, Laurence (2014). *AntConc (Version 3.4.3).* Tokyo.

Archer, Dawn (2005). *Questions and answers in the English courtroom (1640-1760)*: *A sociopragmatic analysis.* Amsterdam: Benjamins.

Archer, Dawn (2013). "Historical Pragmatics: Evidence From The Old Bailey." *Transactions of the Philological Society* 112 (2): 259–277.

Aronsson, K. L. Jonsson and P. Linell (1987). "The Courtroom Hearing as a Middle Ground: Speech Accommodation by Lawyers and Defendants." *Journal of Language and Social Psychology* 6 (2): 99–115.

Ash, Sharon (2013). "Social class." Eds. Jack K. Chambers and Natalie Schilling, *The Handbook of Language Variation and Change*, 2nd edn. Hoboken: Wiley. 350–367.

Auer, Anita (2012). "Late Modern English: Standardization." Eds. Alexander Bergs and Laurel J. Brinton, *English historical linguistics*: *An international handbook*. Berlin: De Gruyter Mouton. 939–952.

Auer, Anita and Victorina González-Díaz (2005). "Eighteenth-century prescriptivism in English: A re-evaluation of its effects on actual language usage." *Multilingua - Journal of Cross-Cultural and Interlanguage Communication* 24 (4): 317–341.

Auer, Anita and Anja Voeste (2012). "Grammatical variables." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 253–270.

Bailey, Richard W. (1996). *Nineteenth-century English.* Ann Arbor, Mich: University of Michigan Press.

Baker, Paul (2010). *Sociolinguistics and corpus linguistics.* Edinburgh: Edinburgh University Press.

Bax, Randy C. (2000). "A Network Strength Scale for the Study of Eighteenth-Century English." *European Journal of English Studies* 4 (3): 277–289.

Beal, Joan C. (2004). *English in modern times 1700 - 1945.* London: Hodder Education.

Beal, Joan C. (2010). "Prescriptivism and the suppression of variation." Ed. Raymond Hickey, *Eighteenth-century English*: *Ideology and change*. Cambridge: Cambridge University Press. 21–37.

Beal, Joan C. (2012a). "Can't see the wood for the trees? Corpora and the study of Late Modern English." Eds. Manfred Markus, Yoko Iyeiri, Reinhard Heuberger and Emil Chamson, *Middle and modern English corpus linguistics*: *A multi-dimensional approach*. Amsterdam: Benjamins. 13–30.

Beal, Joan C. (2012b). "Periods: Late Modern English." Eds. Alexander Bergs and Laurel J. Brinton, *English historical linguistics*: *An international handbook*. Berlin: De Gruyter Mouton. 63–78.

Beard, John R. (1854). *Cassell's Lessons in English; Containing a Practical Grammar, Adapted for the Use of the Self-Educating Student.* London: John Cassell.

Bell, Allan (1984). "Language style as audience design." *Language in Society* 13 (2): 145–205.

Bell, Allan (2001). "Back in style: reworking audience design." Eds. Penelope Eckert and John R. Rickford, *Style and sociolinguistic variation*. Cambridge: Cambridge University Press. 139–169.

Berger, Dieter A. (1978). *Die Konversationskunst in England 1660-1740.* München: Wilhelm Fink.

Bergs, Alexander (2005). *Social networks and historical sociolinguistics*: *Studies in morphosyntactic variation in the Paston letters (1421 - 1503).* Berlin: Mouton de Gruyter.

Bergs, Alexander (2012). "The Uniformitarian Principle and the risk of anachronisms in language and social history." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 80–98.

Biber, Douglas (1988). *Variation across speech and writing.* Cambridge: Cambridge University Press.

Biber, Douglas (1989). "A typology of English texts." *Linguistics* 27 (1): 3–44.

Biber, Douglas (2001). "Dimensions of Variation among 18th-Century Speech-Based and Written Registers." Eds. Hans-Jürgen Diller and Manfred Görlach, *Towards a history of English as a history of genres*. Heidelberg: Winter. 89–109.

Biber, Douglas (2004a). "Historical patterns for the grammatical marking of stance: A cross-register comparison." *Journal of Historical Pragmatics* 5 (1): 107–136.

Biber, Douglas (2004b). "Modal use across registers and time." Eds. Anne Curzan and Kimberly Emmons, *Studies in the history of the English language II*: *Unfolding conversations*. Berlin: Mouton de Gruyter. 189–216.

Biber, Douglas, Susan Conrad and Randi Reppen (1998). *Corpus linguistics*: *Investigating language structure and use.* Cambridge: Cambridge University Press.

Biber, Douglas and Edward Finegan (1987). "Historical drift in three English genres." Eds. Diana McCarthy and Geoffrey Sampson (2005), *Corpus linguistics*: *Readings in a widening discipline*. London: Continuum. 67–77.

Biber, Douglas and Edward Finegan (1989). "Drift and the Evolution of English Style: A History of Three Genres." *Language* 65 (3): 487–517.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan (1999). *Longman grammar of spoken and written English.* Harlow: Longman.

Bock, J. Kathryn (1986). "Syntactic persistence in language production." *Cognitive Psychology* 18 (3): 355–387.

Bowie, Jill, Sean Wallis and Bas Aarts (2013). "Contemporary change in modal usage in spoken British English: mapping the impact of 'genre'." Eds. Juana I. Marín Arrese, Marta Carretero, Jorge Arús Hita and Johan van der Auwera, *English modality*: *Core, periphery and evidentiality*. Berlin: de Gruyter. 57–94.

Breheny, Patrick and Woodrow Burchett (2014). *R Package visreg, version 2.1-1*: *Visualization of Regression Models.*

Breivik, Leiv E. and Ana E. Martínez-Insua (2008). "Grammaticalization, Subjectification and Non-Concord in English Existential Sentences." *English Studies* 89 (3): 351–362.

Brinton, Laurel J. (1991). "The origin and development of quasimodal *have to* in English. Paper presented at the Workshop on "The Origin and Development of Verbal Periphrases", 10th International Conference on Historical Linguistics (ICHL 10) Amsterdam, August 16, 1991. 1–31. <http://faculty.arts.ubc.ca/lbrinton/HAVETO.PDF> Accesed 5 December 2017.

Brinton, Laurel J., Stefan Dollinger and Margery Fee (2012). "Balanced corpora and quotation databases: Taking shortcuts or expanding methodological scope?" Eds. Jukka Tyrkkö, Matti Kilpiö, Terttu Nevalainen and Matti Rissanen, *Outposts of Historical Corpus Linguistics: From the Helsinki Corpus to a Proliferation of Resources.* n.p. <http://www.helsinki.fi/varieng/series/volumes/10/brinton_dollinger_fee>. Accessed 15 June 2015.

Britain, David (2002). “Diffusion, levelling, simplification and reallocation in past tense BE in the English Fens.” *Journal of Sociolinguistics* 6 (1): 16–43.

Britain, David (2012). “Innovation diffusion in sociohistorical linguistics.” Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 451–464.

Britain, David and Andrea Sudbury (2002). “*There’s sheep and there’s penguins*: convergence, ‘drift’ and ‘slant’ in New Zealand and Falkland Island English.” Eds. Mari C. Jones and Edith Esch, *Language change*: *The interplay of internal, external and extra-liguistic factors*. Berlin: Mouton de Gruyter. 209–240.

Bullen, Henry S. J. and Charles Heycock (1853). *Linguae Anglicanae Clavis; or Rudiments of English Grammar, So Arranged for the Use of Schools, as to Form a New and Easy Introduction to Latin and other Classical Grammars.* London: Arthur Hall, Virtue & Co.

Burke, Peter (2004). *Languages and communities in early modern Europe.* Cambridge: Cambridge University Press.

Butler, Judith (1990). *Gender Trouble*: *Feminism and the Subversion of Identity.* Hoboken: Taylor and Francis.

Bybee, Joan L. Revere D. Perkins and William Pagliuca (1994). *The evolution of grammar*: *Tense, aspect, and modality in the languages of the world.* Chicago: University of Chicago Press.

Cameron, Deborah (2008). “Issues of Gender in Modern English.” Eds. Haruko Momma and Michael Matto, *A companion to the history of the English language*. Malden, Mass: Wiley-Blackwell. 293–302.

Canadine, David, ed. (2016). *Oxford Dictionary of National Biography. Online edition.* <http://www.oxforddnb.com/>. Accessed 4 July 2016.

Carter, Ronald and Michael McCarthy (2006). *Cambridge grammar of English.* Cambridge: Cambridge University Press.

Cecconi, Elisabetta (2012). *The language of defendants in the 17th-century English courtroom*: *A socio-pragmatic analysis of the prisoners’ interactional role and representation.* Bern: Lang.

Chambers, Jack K. (1995). *Sociolinguistic theory*: *Linguistic variation and its social significance.* Oxford: Blackwell.

Chambers, Jack K. (2009). *Sociolinguistic theory*: *Linguistic variation and its social significance.* Oxford: Wiley-Blackwell.

Chambers, Jack K. (2013). “Patterns of variation including change.” Eds. Jack K. Chambers and Natalie Schilling, *The Handbook of Language Variation and Change*, 2nd edn. Hoboken: Wiley. 297–323.

Chapman, Carol (1998). “A subject-verb agreement hierarchy.” Eds. Richard M. Hogg and Linda van Bergen, *Historical Linguistics 1995, Volume 2: Germanic linguistics*: *Selected papers from the 12th International Conference on Historical Linguistics, Manchester, August 1995*. Amsterdam: Benjamins. 35-44.

Cheshire, Jenny (1999). “Spoken standard English.” Eds. Tony Bex and Richard J. Watts, *Standard English*: *The widening debate*. London: Routledge. 129–148.

Cheshire, Jenny (2002). “Sex and gender in variationist research.” Eds. Jack K. Chambers, Peter Trudgill and Natalie Schilling-Estes, *The handbook of language variation and change*. Malden, Mass: Blackwell. 423–443.

Cheshire, Jenny and Sue Fox (2009). "*Was/were* variation: A perspective from London." *Language Variation and Change* 21 (1): 1–38.
Ching, Marvin K. L. (2001). "Plural *you/y'all* variation by a court judge: situational use." *American Speech* 76 (2): 115–127.
City Lands Committee (1778). *Journals.* Vol. 70.
Claridge, Claudia (2012). "Linguistic Levels: Styles, registers, genres, text types." Eds. Alexander Bergs and Laurel J. Brinton, *English historical linguistics*: *An international handbook*. Berlin: De Gruyter Mouton. 237–253.
Close, Joanne and Bas Aarts (2010). "Current Change in the Modal System of English: A Case Study of *Must, Have to,* and *Have Got To*." Eds. Ursula Lenker, Judith Huber and Robert Mailhammer, *English Historical Linguistics 2008*, *Volume I: The history of English verbal and nominal constructions*. Amsterdam: Benjamins. 165–181.
Coates, Jennifer (1983). *The semantics of the modal auxiliaries.* London: Croom Helm.
Collins, Peter (2009). *Modals and quasi-modals in English.* Amsterdam: Rodopi.
Collins, Peter (2012). "Singular agreement in *there*-existentials: An intervarietal corpus-based study." *English World-Wide* 33 (1): 53–68.
Cornips, Leonie E. A. and Karen P. Corrigan (2005). "Toward an integrated approach to syntactic variation: A retrospective and prospective synopsis." Eds. Leonie E. A. Cornips and Karen P. Corrigan, *Syntax and variation*: *Reconciling the biological and the social*. Amsterdam: Benjamins. 1–27.
Court of Aldermen (1725). *Repertories of the Court of Aldermen, London.* Vol. 129: 29 September and 7 October 1725. London Metropolitan Archives.
Crane, George (1843). *The Principles of Language; Exemplified in a Practical English Grammar. With Copious Exercises. Designed as an Introduction to the Study of Languages Generally, for the Use of Schools, and Self-Instruction.* London: Whittaker and Co.
Crawford, William J. (2005). "Verb Agreement and Disagreement: A Corpus Investigation of Concord Variation in Existential *There + Be* Constructions." *Journal of English Linguistics* 33 (1): 35–61.
Cressy, David (1980). *Literacy and the social order*: *Reading and writing in Tudor and Stuart England.* Cambridge: Cambridge University Press.
Crombie, Alexander (1809). *A Treatise on the Etymology and Syntax of the English Language* [1802], 2nd edn. London: J. Johnson.
Crystal, David (2008). *A dictionary of linguistics and phonetics* [1980], 6th edn. Malden, Mass: Blackwell.
Culpeper, Jonathan and Dawn Archer (2007). *Sociopragmatic Corpus (SCP). A specialised subsection of A Corpus of English Dialogues (1560-1760).*
Culpeper, Jonathan and Merja Kytö (2000). "Gender voices in the spoken interaction of the past: a pilot study based on Early Modern English trial proceedings." Eds. Dieter Kastovsky and Arthur Mettinger, *The history of English in a social context*: *A contribution to historical sociolinguistics*. Berlin: Mouton de Gruyter. 53–90.
Culpeper, Jonathan and Merja Kytö (2010). *Early modern English dialogues*: *Spoken interaction as writing.* Cambridge: Cambridge University Press.
Curtis, John C. (1876). *An English Grammar for Schools.* London: Simpkin, Marshall, & Co.

Curzan, Anne (2012). “Interdisciplinarity and Historiography: Periodization in the history of the English language.” Eds. Alexander Bergs and Laurel J. Brinton, *English historical linguistics*: *An international handbook*. Berlin: De Gruyter Mouton. 1233–1256.

D’Arcy, Alexandra (2017). *Discourse-pragmatic variation in context*: *Eight hundred years of like.* Amsterdam: Benjamins.

Dawnay, William H. (1857). *An Elementary English Grammar.* London: Longman, Brown, Green, Longmans, and Roberts.

de Haan, Ferdinand (2012). “The Relevance of Constructions for the Interpretation of Modal Meaning: The Case of *Must*.” *English Studies* 93 (6): 700–728.

de Smet, Hendrik (2005). “A corpus of Late Modern English texts.” *ICAME Journal* 29: 69–82.

Denison, David (1993). *English historical syntax*: *Verbal constructions.* London: Longman.

Denison, David (1998). “Syntax.” Eds. Suzanne Romaine and Richard M. Hogg, *The Cambridge history of the English language, Volume IV: 1776-1997*. Cambridge: Cambridge University Press. 92–329.

Depraetere, Ilse and Susan Reed (2006). “Mood and modality in English.” Eds. Bas Aarts and April McMahon, *The handbook of English linguistics*. Malden, Mass: Blackwell. 269–290.

Depraetere, Ilse and An Verhulst (2008). “Source of modality: a reassessment.” *English Language and Linguistics* 12 (1). 1–25.

Deumert, Ana (2003). “Bringing speakers back in? Epistemological reflections on speaker-oriented explanations of language change.” *Language Sciences* 25 (1): 15–76.

Devereaux, Simon (1996). “The City and the Sessions Paper: ‘Public Justice’ in London, 1770-1800.” *Journal of British Studies* 35 (4): 466–503.

Devereaux, Simon (2007). “From sessions to newspaper? Criminal trial reporting, the nature of crime, and the London press, 1770-1800.” *London Journal* 32 (1): 1–27.

Diller, Hans-Jürgen, Hendrik de Smet and Jukka Tyrkkö (2011). *The Corpus of Late Modern English Texts, version 3.0.* <https://perswww.kuleuven.be/~u0044428/clmet3_0.htm>.

Dittmar, Norbert (1997). *Grundlagen der Soziolinguistik*: *Ein Arbeitsbuch mit Aufgaben.* Tübingen: Max Niemeyer.

Dollinger, Stefan (2006). “The Modal Auxiliaries *have to* and *must* in the Corpus of Early Ontario English: Gradient Change and Colonial Lag.” *The Canadian Journal of Linguistics / La revue canadienne de linguistique* 51 (2): 287–308.

Dossena, Marina (2012). “The study of correspondence: Theoretical and methodological issues.” Eds. Marina Dossena and Gabriella Del Lungo Camiciotti, *Letter Writing in Late Modern Europe*. Amsterdam: Benjamins. 13–30.

Dossena, Marina and Ingrid Tieken-Boon van Ostade, eds. (2008). *Studies in late modern English correspondence*: *Methodology and data.* Bern: Lang.

Durrell, Martin (2015). “‘Representativeness’, ‘Bad Data’, and legitimate expectations: What can an electronic historical corpus tell us that we didn’t actually know already (and how)?” Eds. Jost Gippert and Ralf Gehrke, *Historical corpora*: *Challenges and perspectives*. Tübingen: Narr. 13–34.

Eckert, Penelope (1989). “The whole woman: Sex and gender differences in variation.” *Language Variation and Change* 1 (3): 245–267.

Elspaß, Stephan (2012). "The Use of Private Letters and Diaries in Sociolinguistic Investigation." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 156–169.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015a). "About the Proceedings - Publishing history of the Proceedings." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <https://www.oldbaileyonline.org/static/Publishinghistory.jsp>. Accessed 12 March 2017.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015b). "About the Proceedings - The value of the Proceedings as a historical source." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <http://www.oldbaileyonline.org/static/Value.jsp>. Accessed 28 March 2017.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015c). "Communities - Homosexuality." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <http://www.oldbaileyonline.org/static/Gay.jsp>. Accessed 16 June 2016.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015d). "Crime, Justice and Punishment - Crimes tried at the Old Bailey." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <http://www.oldbaileyonline.org/static/Crimes.jsp>. Accessed 29 September 2015.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015e). "Crime, Justice and Punishment - Judges and Juries." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <https://www.oldbaileyonline.org/static/Judges-and-juries.jsp>. Accessed 20 July 2016.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015f). "Crime, Justice and Punishment - Trial Procedures." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <http://www.oldbaileyonline.org/static/Trial-procedures.jsp>. Accessed 10 May 2016.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015g). "Gender in the Proceedings." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <http://www.oldbaileyonline.org/static/Gender.jsp>. Accessed 12 June 2016.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015h). "History of The Old Bailey Courthouse." *Old Bailey Proceedings Online (Version 7.2, March 2015)*. <https://www.oldbaileyonline.org/static/The-old-bailey.jsp>. Accessed 1 July 2016.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015i). "Research and Study Guides - How to Read an Old Bailey Trial." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <http://www.oldbaileyonline.org/static/HowToReadTrial.jsp>. Accessed 12 June 2016.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015j). "Statistics Search." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <https://www.oldbaileyonline.org/forms/formStats.jsp>. Accessed 12 June 2016.

Emsley, Clive, Tim Hitchcock and Robert Shoemaker (2015k). "The Value of the Proceedings as a Historical Source." *Old Bailey Proceedings Online (Version 7.2, March 2015).* <http://www.oldbaileyonline.org/static/Value.jsp>. Accessed 13 June 2016.

Fairclough, Norman (1992). *Discourse and social change.* Cambridge: Polity Press.

Film Study Center at Harvard University (2000). *Martha Ballard's Diary Online.* Transcribed by Robert R. McCausland and Cynthia MacAlman McCausland. <http://dohistory.org/diary/index.html>. Accessed 25 July 2016.

Finegan, Edward (1992). “Style and standardization in England: 1700-1900.” Eds. Tim W. Machan and Charles T. Scott, *English in its social contexts*: *Essays in historical sociolinguistics*. Oxford: Oxford University Press. 103–130.

Finegan, Edward (2012). “Standardization: Prescriptive tradition.” Eds. Alexander Bergs and Laurel J. Brinton, *English historical linguistics*: *An international handbook*. Berlin: De Gruyter Mouton. 967–980.

Finegan, Edward and Douglas Biber (2001). “Register variation and social dialect variation: the Register Axiom.” Eds. Penelope Eckert and John R. Rickford, *Style and sociolinguistic variation*. Cambridge: Cambridge University Press. 235–267.

Fischer, Olga (1994). “The development of quasi-auxiliaries in English and changes in word order.” *Neophilologus* 78 (1): 137–164.

Fischer, Olga (2008). “History of English syntax.” Eds. Haruko Momma and Michael Matto, *A companion to the history of the English language*. Malden, Mass: Wiley-Blackwell. 57–68.

Fitzmaurice, Susan (2004). “Orality, Standardization, and the Effects of Print Publication on the Look of Standard English in the 18th Century.” Eds. Marina Dossena and Roger Lass, *Methods and data in English historical dialectology*. Bern: Lang. 351–383.

Fitzmaurice, Susan M. (2000). “The Spectator, the Politics of Social Networks, and Language Standardisation in Eighteenth Century England.” Ed. Laura Wright, *The Development of Standard English, 1300-1800: Theories, Descriptions, Conflicts*. Cambridge: Cambridge University Press.

Fitzmaurice, Susan M. (2002). “Politeness and modal meaning in the construction of humiliative discourse in an early eighteenth-century network of patron–client relationships.” *English Language and Linguistics* 6 (2): 239–265

Fleischman, Suzanne (1990). *Tense and narrativity*: *From medieval performance to modern fiction.* London: Routledge.

Fludernik, Monika (1991). “The historical present tense yet again: Tense switching and narrative dynamics in oral and quasi-oral storytelling.” *Text - Interdisciplinary Journal for the Study of Discourse* 11 (3): 365–397.

Fox, John (2003). “Effect Displays in R for Generalised Linear Models.” *Journal of Statistical Software* 8 (15): 1–9.

Fox, John and Sanford Weisberg (2015). *R Package car, version 2.0-25*: *Companion to Applied Regression.*

Fox, John, Sanford Weisberg, Michael Friendly, Jangman Hong, Robert Andersen, David Firth and Steve Taylor (2014). *R Package effects, version 3.0-0*: *Effect Displays for Linear, Generalized Linear,Multinomial-Logit, Proportional-Odds Logit Models and Mixed-Effects Models.*

Fritz, Clemens W. A. (2007). *From English in Australia to Australian English: 1788-1900.* Frankfurt am Main: Lang.

Furmaniak, Grégory (2011). “On the emergence of the epistemic use of *must*.” *SKY Journal of Linguistics* 24: 41–73.

Gardiner, Jane (1799). *The young ladies’ English grammar; adapted to the differential classes of learners.* High-Ousegate: Printed by Thomas Wilson and Robert Spencer.

Gatrell, V. A. C. (1990). “Crime, authority and the policeman-state.” Ed. Francis M. L. Thompson, *The Cambridge social history of Britain 1750 - 1950. Vol. 3: Social agencies and institutions*. Cambridge: Cambridge University Press. 243–310.

Goffman, Erving (1981). *Forms of talk.* Philadelphia: University of Pennsylvania Press.
Görlach, Manfred (1991). *Introduction to Early Modern English.* Cambridge: Cambridge University Press.
Görlach, Manfred (1998). *An annotated bibliography of nineteenth-century grammars of English.* Amsterdam: Benjamins.
Görlach, Manfred (1999). *English in nineteenth-century England*: *An introduction.* Cambridge: Cambridge University Press.
Görlach, Manfred (2001). *Eighteenth-century English.* Heidelberg: Winter.
Gurney, Thomas (1752). *Brachygraphy: Or Short-writing,* 2nd edn. London: [no publisher].
Halas, Ana (2012). "The Change in the Form of the Present Perfect in Middle English." Eds. Vesna Lopicic and Biljana M. Ilic, *Challenging Change*: *Literary and Linguistic Responses*. Newcastle upon Tyne: Cambridge Scholars Publishing. 223–234.
Harrell Jr. Frank E. (2014). *R Package rms, version 4.2-0*: *Regression Modeling Strategies.*
Haugland, Kari E. (1995). "Is't allow'd or ain't it? On contraction in early grammars and spelling books." *Studia Neophilologica* 67 (2): 165–184.
Hay, Jennifer and Daniel Schreier (2004). "Reversing the trajectory of language change: Subject–verb agreement with *be* in New Zealand English." *Language Variation and Change* 16 (3): 209–235.
Hernández Campoy, Juan M. and Natalie Schilling (2012). "The application of the quantitative paradigm to historical sociolinguistics: problems with the Generalizability Principle." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 63–79.
Hickey, Raymond (2010a). "Attitudes and concerns in eighteenth-century English." Ed. Raymond Hickey, *Eighteenth-century English*: *Ideology and change*. Cambridge: Cambridge University Press. 1–20.
Hickey, Raymond, ed. (2010b). *Eighteenth-century English*: *Ideology and change.* Cambridge: Cambridge University Press.
Hickey, Raymond (2012). "Internally- and externally-motivated language change." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 387–407.
Higginson, Edward (1864). *An English Grammar Specially Intended for Classical Schools and Private Students.* London: Longman, Green, Longman, Roberts, & Green.
Hiley, Richard (1853). *English Grammar, and Style; to which is Added, Advice to the Student, on the Attainment and Application of Knowledge (Fifth edition, considerably improved, and stereotyped)* [1835]. London: Longman, Brown, Green, and Longmans.
Hitchcock, Tim, Robert Shoemaker, Clive Emsley, Sharon Howard and Jamie McLaughlin (2015). *The Old Bailey Proceedings Online, 1674-1913. Version 7.2.* <www.oldbaileyonline.org>.
Holmes, Janet (1994). "Inferring language change from computer corpora: Some methodological problems." *ICAME Journal* 18: 27–40.
Hope, Jonathan (1993). "Second Person Singular Pronouns in Records of Early Modern 'Spoken' English." *Neuphilologische Mitteilungen* 94 (1): 83–100.
Hort, William J. (1822). *An Introduction to English Grammar: Equally Adapted to Domestic and to School Education.* London: Longman, Hurst, Rees, Orme, and Brown.

Howell, T. B. (1816). *A complete collection of state trials and proceedings for high treason and other crimes and misdemeanors from the earliest period to the year 1783*: *With notes and other illustrations. Vol. XIX.* London: T. C. Hansard.

Huber, Magnus (2007). "The Old Bailey Proceedings, 1674–1834: Evaluating and annotating a corpus of 18th- and 19th-century spoken English." *Annotating Variation and Change: Studies in Variation, Contacts and Change in English* 1: 1–39.

Huber, Magnus (2010). "Trial proceedings as a source of spoken English. A critical evaluation based on negative contraction in the Proceedings of the Old Bailey, 1674-1913." Eds. Jörg Helbig and René Schallegger, A*nglistentag 2009 Klagenfurt. Proceedings.* Trier: Wissenschaftlicher Verlag Trier. 65–78.

Huber, Magnus (2017). "Structural and sociolinguistic factors conditioning the choice of relativizers in 18th and 19th century English: A diachronic study based on the Old Bailey Corpus." *Nordic Journal of English Studies* 16 (1): 74-119.

Huber, Magnus, Magnus Nissel, Patrick Maiwald and Bianca Widlitzki (2012). *The Old Bailey Corpus 1.0. Spoken English in the 18th and 19th centuries*. <www.uni-giessen.de/oldbaileycorpus>.

Huddleston, Rodney D. and Geoffrey K. Pullum (2002). *The Cambridge grammar of the English language.* Cambridge: Cambridge University Press.

Hudson, Richard (1999). "Subject–verb agreement in English." *English Language and Linguistics* 3 (2): 173–207.

Hundt, Marianne (2014a). "Introduction: Late Modern English syntax in its linguistic and socio-historical context." Ed. Marianne Hundt, *Late Modern English syntax*. Cambridge: Cambridge University Press. 1–10.

Hundt, Marianne, ed. (2014b). *Late Modern English syntax.* Cambridge: Cambridge University Press.

Hundt, Marianne and Christian Mair (1999). "'Agile' and 'Uptight' Genres: The Corpus-based Approach to Language Change in Progress." *International Journal of Corpus Linguistics* 4 (2): 221–242.

Hutchins, Joseph (1791). *An abstract of the first principles of English grammar. Compiled for the use of his own school.* Philadelphia: Printed by Thomas Dobson.

Ihalainen, Ossi (1994). "The dialects of England since 1776." Ed. Robert Burchfield, *The Cambridge history of the English language, Volume V: English in Britain and Overseas: Origins and Development*. Cambridge: Cambridge University Press. 197–274.

Jacobsson, Bengt (1979). "Modality and the modals of necessity *must* and *have to*." *English Studies* 60 (3): 296–312.

James, J. H. (1847). *The Elements of Grammar, According to Dr. Becker's System, Displayed by the Structure of the English Tongue, (With Copious Examples from the Best Writers,) Arranged as a Practice for Translation into Foreign Languages.* London: Longman, Brown, Green, and Longmans.

Jankowski, Bridget (2004). "A transatlantic perspective of variation and change in English deontic modality." *Toronto Working Papers in Linguistics* 23 (2): 85–113.

Jespersen, Otto (1961). *A modern English grammar on historical principles*: *Part 4: Syntax (Third Volume).* London: Allen and Unwin.

Johansson, Stig (2013). "Modals and semi-modals of obligation in American English: some aspects of developments from 1990 until the present day." Eds. Bas Aarts, Joanne Close, Geoffrey Leech and Sean Wallis, *The verb phrase in English*: *Investigating recent language change with corpora*. Cambridge: Cambridge University Press. 372–380.

Johnstone, Barbara (1987). "'He says … so I said': verb tense alternation and narrative depictions of authority in American English." *Linguistics* 25 (1): 33–52.

Kennedy, Graeme (2002). "Variation in the Distribution of Modal Verbs in the British National Corpus." Eds. Randi Reppen, Susan M. Fitzmaurice and Douglas Biber, *Using corpora to explore linguistic variation*. Amsterdam: Benjamins. 73–90.

Kiełkiewicz-Janowiak, Agnieszka (2002). *'Women's language'? A socio-historical view: private writings in early New England.* Poznań: Motivex.

Kiełkiewicz-Janowiak, Agnieszka (2012). "Class, age, and gender-based patterns." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 307–331.

Kiesling, Scott F. (2013). "Constructing identity." Eds. Jack K. Chambers and Natalie Schilling, *The Handbook of Language Variation and Change*, 2nd edn. Hoboken: Wiley. 448–467.

Klemola, Juhani (1996). *Non-standard periphrastic do: A study of variation and change*. PhD dissertation. Essex.

Koch, Peter and Wulf Oesterreicher (1985). "Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte." *Romanistisches Jahrbuch* 36: 15–43.

Kortmann, Bernd and Kerstin Lunkenheimer, eds. (2013). *The Electronic World Atlas of Varieties of English.* Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://ewave-atlas.org/>. Accessed 25 July 2016.

Krug, Manfred G. (2000). *Emerging English modals*: *A corpus-based study of grammaticalization.* Berlin: Mouton de Gruyter.

Kytö, Merja (1994). "*Be* vs. *have* with intransitives in Early Modern English." Eds. Francisco Fernández, Miguel Fuster and Juan J. Calvo, *English historical linguistics 1992*: *Papers from the 7th International Conference on English Historical Linguistics, Valencia, 22 - 26 September 1992*. Amsterdam: Benjamins. 179–190.

Kytö, Merja (1997). "*Be/have* + past participle: The choice of the auxiliary with intransitives from Late Middle to Modern English." Eds. Matti Rissanen, Merja Kytö and Kirsi Heikkonen, *English in transition*: *Corpus-based studies in linguistic variation and genre styles*. Berlin: Mouton de Gruyter. 17–86.

Kytö, Merja (2011). "Corpora and historical linguistics." *Revista Brasileira de Linguística Aplicada* 11: 417–457.

Kytö, Merja (2012). "New Perspectives, Theories and Methods: Corpus linguistics." Eds. Alexander Bergs and Laurel J. Brinton, *English historical linguistics*: *An international handbook*. Berlin: De Gruyter Mouton. 1509–1531.

Kytö, Merja, Peter Grund and Terry Walker (2011). *Testifying to language and life in early modern England.* Amsterdam: Benjamins.

Kytö, Merja and Matti Rissanen (1983). "The syntactic study of American English: The variationist at the mercy of his corpus?" *Neuphilologische Mitteilungen* 84 (4): 470–490.

Kytö, Merja, Juhani Rudanko and Erik Smitterberg (2000). "Building a bridge between the present and the past: A corpus of 19th-century English." *ICAME Journal* (24): 85–97.

Kytö, Merja and Erik Smitterberg (2006). "19th-century English: An Age of Stability or a Period of Change?" Eds. Roberta Facchinetti and Matti Rissanen, *Corpus-based Studies of Diachronic English*. Bern: Lang. 199–230.

Kytö, Merja and Terry Walker (2003). "The Linguistic Study of Early Modern English Speech-Related Texts: How 'Bad' can 'Bad' Data Be?" *Journal of English Linguistics* 31 (3): 221–248.

Labov, William (1972a). *Language in the inner city*: *Studies in the Black English vernacular.* Philadelphia: University of Pennsylvania Press.

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

Labov, William (1990). "The intersection of sex and social class in the course of linguistic change." *Language Variation and Change* 2 (2): 205–254.

Labov, William (1994). *Principles of linguistic change*: *Internal Factors.* Oxford: Blackwell.

Labov, William (2001). *Principles of linguistic change*: *Social factors.* Malden, Mass: Blackwell.

Labov, William (2006). *The social stratification of English in New York City* [1966], 2nd edn. Cambridge: Cambridge University Press.

Labov, William (2010). *Principles of linguistic change*: *Cognitive and cultural factors.* Malden, Mass: Wiley-Blackwell.

Laitinen, Mikko (2009). "Singular YOU WAS/WERE Variation and English Normative Grammars in the Eighteenth Century." Eds. Arja Nurmi, Minna Nevala and Minna Palander-Collin, *The language of daily life in England (1400-1800)*. Amsterdam: Benjamins. 199–217.

Langbein, John H. (2003). *The origins of adversary criminal trial.* Oxford: Oxford University Press.

Langer, Nils and Agnete Nesse (2012). "Linguistic purism." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 607–625.

Lass, Roger (1997). *Historical linguistics and language change.* Cambridge: Cambridge University Press.

Lavandera, Beatriz R. (1978). "Where does the sociolinguistic variable stop?" *Language in Society* 7 (2): 171-182.

Leech, Geoffrey (2003). "Modality on the move: The English modal auxiliaries 1961-1992." Eds. Roberta Facchinetti, Manfred G. Krug and Frank Palmer, *Modality in contemporary English*. Berlin: Mouton de Gruyter. 223–240.

Leech, Geoffrey (2011). "The modals ARE declining: Reply to Neil Millar's 'Modal verbs in TIME: Frequency changes 1923–2006', *International Journal of Corpus Linguistics* 14:2 (2009), 191–220." *International Journal of Corpus Linguistics* 16 (4): 547–564.

Leech, Geoffrey (2013). "Where have all the modals gone? An essay on the declining frequency of core modal auxiliaries in recent standard English." Eds. Juana I. Marín Arrese, Marta Carretero, Jorge Arús Hita and Johan van der Auwera, *English modality*: *Core, periphery and evidentiality*. Berlin: de Gruyter. 95–115.

Leech, Geoffrey and Jennifer Coates (1980). "Semantic Indeterminacy and the Modals." Eds. Sidney Greenbaum, Geoffrey N. Leech and Jan Svartvik, *Studies in English linguistics for Randolph Quirk*. London: Longman. 79–90.

Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith (2009). *Change in contemporary English*: *A grammatical study.* Cambridge: Cambridge University Press.
Leech, Geoffrey and Nicholas Smith (2006). "Recent Grammatical Change in Written English 1961-1992: Some Preliminary Findings of a Comparison of American with British English." Eds. Antoinette Renouf and Andrew Kehoe, *The changing face of corpus linguistics*. Amsterdam: Rodopi. 185–204.
Leech, Geoffrey and Nicholas Smith (2009). "Change and constancy in linguistic change: How grammatical usage in written English evolved in the period 1931-1991." Eds. Antoinette Renouf and Andrew Kehoe, *Corpus linguistics: refinements and reassessements*. Amsterdam: Rodopi. 185–204.
Leech, Geoffrey N. and Mick Short (2007). *Style in fiction*: *A linguistic introduction to English fictional prose* [1981], 2nd edn. Harlow: Longman.
Levshina, Natalia (2015). *How to do linguistics with R*: *Data exploration and statistical analysis.* Amsterdam: Benjamins.
Lightfoot, David (1979). *Principles of diachronic syntax.* Cambridge: Cambridge University Press.
Lowth, Robert (1762). *A Short Introduction to English Grammar.* London: Millar, Dodsley & Dodsley.
Lyons, John (1977). *Semantics. Vol. 2.* Cambridge: Cambridge University Press.
Mair, Christian (1997). "Parallel Corpora: A Real-Time Approach to the Study of Language Change in Progress." Ed. Magnus Ljung, *Corpus-based studies in English*: *Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17) Stockholm, May 15 - 19, 1996*. Amsterdam: Rodopi. 195–209.
Mair, Christian (2009). "Corpus linguistics meets sociolinguistics: the role of corpus evidence in the study of sociolinguistic variation and change." Eds. Antoinette Renouf and Andrew Kehoe, *Corpus linguistics: refinements and reassessments*. Amsterdam: Rodopi. 7–32.
Martínez-Insua, Ana E. and Javier Pérez Guerra (2006). "'There's Bjørg': On *There*-sentences in the recent history of English." Eds. Leiv E. Breivik, Sandra Halverson and Kari E. Haugland, *"These things write I vnto thee…"*: *Essays in honour of Bjørg Bækken*. Oslo: Novus Press. 189–212.
Mason, Charles P. (1873). *English Grammar; Including the Principles of Grammatical Analysis* [1858], 18th edn. London: Bell & Daldy.
Matthew, H. C. G. (2001). "The Liberal Age (1851-1914)." Ed. Kenneth O. Morgan, *The Oxford history of Britain*. Oxford: Oxford University Press. 518–581.
May, Allyson N. (2003). *The Bar and the Old Bailey, 1750-1850.* Chapel Hill: University Of North Carolina Press.
Mazzon, Gabriella (2004). *A history of English negation.* Harlow: Pearson/Longman.
McCafferty, Kevin (2014). "'(W)ell are you not got over thinking about going to Ireland yet': the BE-perfect in eighteenth- and nineteenth-century Irish English." Ed. Marianne Hundt, *Late Modern English syntax*. Cambridge: Cambridge University Press. 333–351.
McCarthy, Michael (1998). *Spoken language and applied linguistics.* Cambridge: Cambridge University Press.

McColl Millar, Robert (2012). “Social history and the sociology of language.” Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 41–59.

McFadden, Thomas and Artemis Alexiadou (2006). “Auxiliary selection and counterfactuality in the history of English and Germanic.” Eds. Jutta M. Hartmann and László Molnárfi, *Comparative Studies in Germanic Syntax*. Amsterdam: Benjamins. 237–262.

McFadden, Thomas and Artemis Alexiadou (2010). “Perfects, Resultatives, and Auxiliaries in Earlier English.” *Linguistic Inquiry* 41 (3): 389–425.

McIntosh, Carey (2008). “British English in the Long Eighteenth Century (1660-1830).” Eds. Haruko Momma and Michael Matto, *A companion to the history of the English language*. Malden, Mass: Wiley-Blackwell. 228–234.

McIntyre, Dan, Carol Bellard-Thomson, John Heywood, Tony McEnery, Elena Semino and Mick Short (2004). “Investigating the presentation of speech, writing and thought in spoken British English: A corpus-based approach.” *ICAME Journal* 28: 49–76.

Meechan, Marjory and Michele Foley (1994). “On resolving disagreement: Linguistic theory and variation – There’s bridges.” *Language Variation and Change* 6 (1): 63–85.

Meyerhoff, Miriam (2001). “Dynamics of differentiation: On social psychology and cases of language variation.” Eds. Nikolas Coupland, Srikant Sarangi and Christopher N. Candlin, *Sociolinguistics and social theory*. Harlow: Longman. 61–87.

Millar, Neil (2009). “Modal verbs in TIME: Frequency changes 1923–2006.” *International Journal of Corpus Linguistics* 14 (2): 191–220.

Milroy, James (1992a). “A social model for the Interpretation of language change.” Eds. Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen and Irma Taavitsainen, *History of Englishes*: *New methods and interpretations in historical linguistics*. Berlin: Mouton de Gruyter. 72–91.

Milroy, James (1992b). *Linguistic variation and change*: *On the historical sociolinguistics of English.* Oxford: Blackwell.

Milroy, James (2001). “Language ideologies and the consequences of standardization.” *Journal of Sociolinguistics* 5 (4): 530–555.

Milroy, Lesley (1980). *Language and social networks.* Oxford: Blackwell.

Milroy, Lesley (2012). “Sociolinguistics and Ideologies in Language History.” Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 571–584.

Mitchell, Bruce (1985). *Old English syntax. Volume 2.* Oxford: Oxford University Press.

Molencki, Rafal (2003). “What must needs be explained about *must needs*.” Ed. David Hart, *English modality in context*: *Diachronic perspectives*. Bern: Lang. 71–88.

Montgomery, Michael (1997). “A tale of two Georges: The language of Irish Indian traders in colonial North America.” Ed. Jeffrey L. Kallen, *Focus on Ireland*. Amsterdam: Benjamins. 227–254.

Moore, Colette (2008). “Reporting direct speech in Early Modern slander depositions.” Eds. Graeme Trousdale and Nikolas Gisborne, *Constructional Approaches to English Grammar*. Berlin: Mouton de Gruyter. 399–416.

Moore, Emma (2010). “Interaction between social category and social practice: explaining *was/were* variation.” *Language Variation and Change* 22 (3): 347–371.

Mugglestone, Lynda (2006). "Introduction: A History of English." Ed. Lynda Mugglestone, *The Oxford History of English*. Oxford: Oxford University Press. 1–8.

Murry, Ann (1778). *Mentoria: or, the young ladies' instructor, in familiar conversations on moral and entertaining subjects.* London: Printed by J. Fry & Co.

Myhill, John (1995). "Change and continuity in the functions of the American English modals." *Linguistics* 33 (2): 157–211.

Nevalainen, Terttu (2006). "Vernacular universals? The case of plural *was* in Early Modern English." Eds. Terttu Nevalainen, Juhani Klemola and Mikko Laitinen, *Types of variation*: *Diachronic, dialectical and typological interfaces*. Amsterdam: Benjamins. 351–370.

Nevalainen, Terttu (2009). "Number Agreement in Existential Constructions: A Sociolinguistics Study of Eighteenth-Century English." Eds. Markku Filppula, Juhani Klemola and Heli Paulasto, *Vernacular universals and language contacts*: *Evidence from varieties of English and beyond*. New York: Routledge. 80–102.

Nevalainen, Terttu (2015). "Descriptive adequacy of the S-curve model in diachronic studies of language change." Ed. Christina Sanchez-Stockhammer, *Can We Predict Linguistic Change?* n.p. <http://www.helsinki.fi/varieng/series/volumes/16/nevalainen> Accessed 16 January 2016.

Nevalainen, Terttu and Helena Raumolin-Brunberg (2003). *Historical sociolinguistics*: *Language change in Tudor and Stuart England.* London: Longman.

Nevalainen, Terttu and Helena Raumolin-Brunberg (2012). "Historical Sociolinguistics: Origins, Motivations, and Paradigms." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 22–40.

Nevalainen, Terttu, Helena Raumolin-Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi and Minna Palander-Collin (1998). *Corpus of Early English Correspondence (CEEC)*.

Nevalainen, Terttu, Helena Raumolin-Brunberg, Samuli Kaislaniemi, Mikko Laitinen, Minna Nevala, Arja Nurmi, Minna Palander-Collin, Tanja Säily and Anni Sairio (n.d.). *Corpus of Early English Correspondence Extension (CEECE)*. Unpublished.

Nevalainen, Terttu, Helena Raumolin-Brunberg and Heikki Mannila (2011). "The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change." *Language Variation and Change* 23 (1): 1–43.

Nurmi, Arja (2013). "'All the rest ye must lade yourself': Deontic modality in sixteenth - century English merchant letters." Eds. Marijke J. van der Wal and Gijsbert Rutten, *Touching the Past*: *Studies in the historical sociolinguistics of ego-documents*. Amsterdam: Benjamins. 165–182.

O'Barr, William M. and Bowman K. Atkins (1980). "'Women's language' or 'powerless language'?" Eds. Sally McConnell-Ginet, Ruth Borker and Nelly Furman, *Women and language in literature and society*. New York: Praeger. 93–110.

Oxford University Press (2017). "OED (Oxford English Dictionary) Online. September 2017 Update. <www.oed.com>. Accessed 5 September 2017.

Palmer, Frank (1990). *Modality and the English modals* [1979], 2nd edn. London: Longman.

Palmer, Frank (2003). "Modality in English: Theoretical, descriptive and typological issues." Eds. Roberta Facchinetti, Manfred G. Krug and Frank Palmer, *Modality in contemporary English*. Berlin: Mouton de Gruyter. 1–17.

Percy, Carol (2012). “Standardization: Codifiers.” Eds. Alexander Bergs and Laurel J. Brinton, *English historical linguistics*: *An international handbook*. Berlin: De Gruyter Mouton. 1006–1020.

Phillipps, K. C. (1970). *Jane Austen’s English.* London: Deutsch.

Pietsch, Lukas (2005). “‘Some do and some doesn’t’: Verbal concord variation in the north of the British Isles.” Eds. Bernd Kortmann, Tanja Herrmann, Lukas Pietsch and Susanne Wagner, *A Comparative Grammar of British English Dialects: Agreement, gender, relative clauses*. Berlin: Mouton de Gruyter. 126–209.

Pinnock, William (1830). *A Comprehensive Grammar of the English Language: with Exercises; Written in a Familiar Style; Accompanied with Questions for Examination, and Notes Critical and Explanatory. Intended for the Use of Schools, and for Private Tuition* [1829]. 2nd. edn. London: Poole & Edwards.

Plank, Frans (1984). “The Modals Story Retold.” *Studies in Language* 8 (3): 305–364.

Pratt, Lynda and David Denison (2000). “The language of the Southey–Coleridge Circle.” *Language Sciences* 22 (3): 401–422.

Price, Richard (1999). *British society, 1680-1880*: *Dynamism, containment, and change.* Cambridge: Cambridge University Press.

Priestley, Joseph (1768). *The rudiments of English grammar, adapted to the use of schools*…. London: T. Becket & P.A. DeHondt (Strand); J. Johnson (Paternoster-Row).

Quaglio, Paulo and Douglas Biber (2006). “The Grammar of Conversation.” Eds. Bas Aarts and April McMahon, *The handbook of English linguistics*. Malden, Mass: Blackwell. 692–723

Queen, Robin (2013). “Gender, Sex, Sexuality, and Sexual Identities.” Eds. Jack K. Chambers and Natalie Schilling, *The Handbook of Language Variation and Change*, 2nd edn. Hoboken: Wiley. 368–387.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985). *A comprehensive grammar of the English language.* London: Longman.

Quirk, Randolph and Charles L. Wrenn (1958). *An Old English grammar* [1955], 2nd edn. London: Methuen.

R Core Team (2016). *R: A language and environment for statistical computing. Version 3.3.1.* Vienna: R Foundation for Statistical Computing.

Raumolin-Brunberg, Helena (2006). “Leaders of linguistic change in early modern England.” Eds. Roberta Facchinetti and Matti Rissanen, *Corpus-based studies of diachronic English*. Bern: Lang. 115–134.

Reichmann, Oskar (1990). “Sprache ohne Leitvarietät vs. Sprache mit Leitvarietät: ein Schlüssel für die nachmittelalterliche Geschichte des Deutschen?“ Ed. Werner Besch, *Deutsche Sprachgeschichte*: *Grundlagen, Methoden, Perspektiven*. Frankfurt am Main: Lang. 141–158.

Riordan, Brian (2007). “*There’s two ways to say it*: Modeling nonprestige *there’s*.” *Corpus Linguistics and Linguistic Theory* 3 (2). 233–279.

Rissanen, Matti (1989). “Three problems connected with the use of diachronic corpora.” *ICAME Journal* 13: 16–19.

Rissanen, Matti (1999). “Syntax.” Ed. Roger Lass, *The Cambridge history of the English language, Volume III: 1476-1776*. Cambridge: Cambridge University Press. 187–331.

Roberge, Paul T. (2012). "The Teleology of Change: Functional and Non-Functional Explanations." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 369–386.

Romaine, Suzanne (1982a). *Socio-historical linguistics*: *its status and methodology.* Cambridge: Cambridge University Press.

Romaine, Suzanne (1982b). "The reconstruction of language in its social context: Methodology for a socio-historical linguistic theory." Ed. Anders Ahlqvist, *Papers from the 5th International Conference on Historical Linguistics*. Amsterdam: Benjamins. 293–303.

Romaine, Suzanne (1988). "Historical sociolinguistics: Problems and methodology." Eds. Ulrich Ammon, Norbert Dittmar and Klaus J. Mattheier, *Sociolinguistics: An International Handbook of the Science of Language and Society*, 2nd edn. Berlin: de Gruyter. 1452–1469.

Romaine, Suzanne (1999). *Communicating gender.* Mahwah, NJ: Erlbaum.

Romaine, Suzanne (2005). "Historical Sociolinguistics: Problems and Methodology." Eds. Peter Trudgill, Klaus J. Mattheier and Norbert Dittmar, *Sociolinguistics: an international handbook of the science of language and society*, 2nd edn. Berlin: Mouton de Gruyter. 1696–1703.

RStudio Team (2015). *RStudio: Integrated Development for R.* Boston: RStudio, Inc.

Rühlemann, Christoph (2007). *Conversation in context*: *A corpus-driven approach.* London: Continuum.

Rühlemann, Christoph (2008). "Conversational grammar – bad grammar? A situation-based description of quotative *I goes* in the BNC." *ICAME Journal* 32: 157–178.

Rushton, William (1869). *Rules and Cautions in English Grammar Founded on the Analysis of Sentences.* London: Longmans, Green, and Co.

Rydén, Mats (1979). *An Introduction to the Historical Study of English Syntax.* Stockholm: Almqvist and Wiksell.

Rydén, Mats (1984). "The study of eighteenth century English syntax." Ed. Jacek Fisiak, *Historical syntax*: *Papers presented for the International Conference on Historical Syntax held at Błażejewko, Poland, 31 March - 3 April 1981*. Berlin: Mouton de Gruyter. 509–520.

Rydén, Mats (1991). "The *be/have* variation with intransitives in its crucial phases." Ed. Dieter Kastovsky, *Historical English Syntax*. Berlin: Mouton de Gruyter. 343–354.

Rydén, Mats and Sverker Brorström (1987). *The* be/have *variation with intransitives in English*: *With special reference to the late modern period.* Stockholm: Almqvist and Wiksell.

Sairio, Anni (2006). "Progressives in the Letters of Elizabeth Montagu and her Circle in 1738-1778." Eds. Christiane Dalton-Puffer, Dieter Kastovsky, Nikolaus Ritt and Herbert Schendl, *Syntax, Style and Grammatical Norms*: *English from 1500 - 2000*. Bern: Lang. 167–189.

Sairio, Anni (2009). *Language and letters of the Bluestocking network. Sociolinguistic Issues in Eighteenth-Century Epistolary English.* Helsinki: Modern Language Society.

Sairio, Anni and Minna Palander-Collin (2012). "The reconstruction of prestige patterns in language history." Eds. Juan M. Hernández Campoy and Juan C. Conde-Silvestre, *The handbook of historical sociolinguistics*. Malden, Mass: Wiley-Blackwell. 626–638.

Sakita, Tomoko I. (2002). "Dialogue-internal and external features representing mental imagery of speaker attitudes." *Text - Interdisciplinary Journal for the Study of Discourse* 22 (1): 83–105.

Sankoff, David (1988). "Sociolinguistics and syntactic variation." Ed. Frederick J. Newmeyer, *Language*: *The socio-cultural context*. Cambridge: Cambridge University Press. 140–161.

Schiffrin, Deborah (1981). "Tense variation in narrative." *Language* 57 (1): 45–62.

Schilling-Estes, Natalie and Walt Wolfram (1994). "Convergent explanation and alternative regularization patterns: *Were/weren't* leveling in a vernacular English variety." *Language Variation and Change* 6 (3): 273–302.

Schneider, Edgar (2002). "Investigating variation and change in written documents." Eds. Jack K. Chambers, Peter Trudgill and Natalie Schilling-Estes, *The handbook of language variation and change*. Malden, Mass: Blackwell. 67–96.

Schneider, Edgar (2013). "Investigating Historical Variation and Change in Written Documents. New Perspectives." Eds. Jack K. Chambers and Natalie Schilling, *The Handbook of Language Variation and Change*, 2nd edn. Hoboken: Wiley. 57–81.

Schulz, Monika Edith (2011). "Possession and obligation." Eds. Nuria Hernández, Daniela Kolbe and Monika Edith Schulz, *A comparative grammar of British English dialects*: *Modals, pronouns and complement clauses*. Berlin: de Gruyter. 19–51.

Shepherd, Susan C. (1982). "From Deontic to Epistemic: An Analysis of Modals in the History of English, Creoles, and Language Acquisition." Ed. Anders Ahlqvist, *Papers from the 5th International Conference on Historical Linguistics*. Amsterdam: Benjamins. 316–323.

Shoemaker, Robert (2008). "The Old Bailey Proceedings and the Representation of Crime and Criminal Justice in Eighteenth-Century London." *Journal of British Studies* (47): 559–580.

Simpson, John and Edmund Weiner, eds. (1989). *Oxford English Dictionary* [1884–1928], 2nd edn. Oxford: Oxford University Press.

Smith, Jennifer and Sali Tagliamonte (1998). "'We were all thegither... I think we was all thegither': *was* regularization in Buckie English." *World Englishes* 17 (2): 105–126.

Smith, Jeremy (1996). *An historical study of English*: *Function, form and change.* London: Routledge.

Smith, John (1755). *The Printer's Grammar.* London: Printed by L. Wayland.

Smith, K. A. (2007). "Language use and auxiliary selection in the perfect." Ed. Raúl Aranovich, *Split Auxiliary Systems*. Amsterdam: Benjamins. 255–270.

Smith, Nicholas (2003). "Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English." Eds. Roberta Facchinetti, Manfred G. Krug and Frank Palmer, *Modality in contemporary English*. Berlin: Mouton de Gruyter. 241–267.

Smitterberg, Erik (2005). *The progressive in 19th-century English*: *A process of integration.* Amsterdam: Rodopi.

Smitterberg, Erik (2008). "The Progressive and Phrasal Verbs: Evidence of Colloquialization in Nineteenth-Century English?" Eds. Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta and Minna Korhonen, *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*. Amsterdam: Benjamins. 269–289.

Smitterberg, Erik (2012). “Late Modern English: Sociolinguistics.” Eds. Alexander Bergs and Laurel J. Brinton, *English historical linguistics*: *An international handbook*. Berlin: De Gruyter Mouton. 952–965.

Stein, Dieter (1994). “Sorting out the variants: Standardization and social factors in the English language 1600-1800.” Eds. Dieter Stein and Ingrid Tieken-Boon van Ostade, *Towards a standard English*: *1600 - 1800*. Berlin: Mouton de Gruyter. 1–18.

Straaijer, Robin (2010). “Prescription or practice? *Be/have* variation with past participles of mutative intransitive verbs in the letters of Joseph Priestley.” Eds. Ursula Lenker, Judith Huber and Robert Mailhammer, *English Historical Linguistics 2008, Volume I: The history of English verbal and nominal constructions*. Amsterdam: Benjamins. 63–78.

Strang, Barbara (1970). *A history of English.* London: Methuen.

Sundby, Bertil, Anne K. Bjørge and Kari E. Haugland (1991). *A dictionary of English normative grammar, 1700 - 1800.* Amsterdam: Benjamins.

Sutherland, Gillian (1990). “Education.” Ed. Francis M. L. Thompson, *The Cambridge social history of Britain 1750 - 1950. Vol. 3: Social agencies and institutions*. Cambridge: Cambridge University Press. 119–170.

Tagliamonte, Sali A. (1998). “*Was/were* variation across the generations: View from the city of York.” *Language Variation and Change* 10 (2): 153–191.

Tagliamonte, Sali A. (2004). “*Have to, gotta, must*: grammaticalization, variation and specialization in English deontic modality.” Eds. Hans Lindquist and Christian Mair, *Corpus approaches to grammaticalization in English*. Philadelphia: Benjamins. 33–56.

Tagliamonte, Sali A. (2006). *Analysing sociolinguistic variation.* Cambridge: Cambridge University Press.

Tagliamonte, Sali A. (2009). “*There Was Universals; Then There Weren’t*: A Comparative Sociolinguistic Perspective on ‘Default Singulars’.” Eds. Markku Filppula, Juhani Klemola and Heli Paulasto, *Vernacular universals and language contacts*: *Evidence from varieties of English and beyond*. New York: Routledge. 103–132.

Tagliamonte, Sali A. (2012). *Variationist sociolinguistics*: *Change, observation, interpretation.* Malden, Mass: Wiley-Blackwell.

Tagliamonte, Sali A. and R. Harald Baayen (2012). “Models, forests, and trees of York English: Was/were variation as a case study for statistical practice.” *Language Variation and Change* 24 (2): 135–178.

Tagliamonte, Sali A. and Jennifer Smith (2006). “Layering, competition and a twist of fate: Deontic modality in dialects of English.” *Diachronica* 23 (2): 341–380.

Tieken-Boon van Ostade, Ingrid (1987). *The auxiliary* do *in eighteenth-century English*: *A sociohistorical-linguistic approach*. Dordrecht: Foris.

Tieken-Boon van Ostade, Ingrid (2000). “Sociohistorical linguistics and the observer’s paradox.” Eds. Dieter Kastovsky and Arthur Mettinger, *The history of English in a social context*: *A contribution to historical sociolinguistics*. Berlin: Mouton de Gruyter. 441–462.

Tieken-Boon van Ostade, Ingrid (2002). “*You was* and eighteenth-century normative grammar.” Ed. Katja Lenz, *Of dyuersitie & chaunge of language*: *Essays presented to Manfred Görlach on the occasion of his 65th birthday*. Heidelberg: Winter. 88–102.

Tieken-Boon van Ostade, Ingrid (2009). *An Introduction to Late Modern English.* Edinburgh: Edinburgh University Press.

Traugott, Elizabeth C. (1972). *A history of English syntax*: *A transformational approach to the history of English sentence structure.* New York: Holt, Rinehart and Winston.

Traugott, Elizabeth C. (1989). "On the Rise of Epistemic Meanings in English: An Example of Subjectification in Semantic Change." *Language* 65 (1): 31–55.

Traugott, Elizabeth C. (2011). "Constructing the audiences of the Old Bailey Trials 1674-1834." Eds. Päivi Pahta and Andreas H. Jucker, *Communicating early English manuscripts*. Cambridge: Cambridge University Press. 69–80.

Trousdale, Graeme (2003). "Simplification and redistribution: An account of modal verb usage in Tyneside English." *English World-Wide* 24 (2): 271–284.

Turner, Brandon (1840). *A New English Grammar; in which the Principles of that Science are Fully Explained, and Adapted to the Comprehension of Young Persons; Containing a Series of Exercises for Parsing, for Oral Correction, and for Writing, with Questions for Examination.* London: Scott, Webster, and Geary.

Ukaji, Masatomo (1992). "'I not say': bridge phenomenon in syntactic change." Eds. Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen and Irma Taavitsainen, *History of Englishes*: *New methods and interpretations in historical linguistics*. Berlin: Mouton de Gruyter. 453-462.

van der Gaaf, Willem (1931). "*Beon* and *habban* connected with an inflected infinitive." *English Studies* 13: 176–188.

van Kemenade, Ans (1992). "Structural factors in the history of English modals." Eds. Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen and Irma Taavitsainen, *History of Englishes*: *New methods and interpretations in historical linguistics*. Berlin: Mouton de Gruyter. 287–309.

van Leeuwen, Marco H. D. and Ineke Maas (2011). *HISCLASS. A historical international social class scheme.* Leuven: Leuven University Press.

van Leeuwen, Marco H. D., Ineke Maas and Andrew Miles (2002). *HISCO*: *Historical international standard classification of occupations.* Leuven: Leuven University Press.

Visser, Frederikus T. (1963-1973). *An historical syntax of the English language (3 vols.).* Leiden: Brill.

Walker, Anne G. (1987). "Linguistic manipulation, power and the legal setting." Ed. Leah Kedar, *Power through discourse*. Norwood: Ablex. 57–80.

Walker, James A. (2007). "'There's bears back there': Plural existentials and vernacular universals in (Quebec) English." *English World-Wide* 28 (2): 147–166.

Ward, Richard M. (2014). *Print culture, crime and justice in eighteenth-century London.* London: Bloomsbury.

Warner, Anthony R. (1993). *English auxiliaries*: *Structure and history.* Cambridge: Cambridge University Press.

Webster, Noah (1784). *A grammatical institute, of the English language,… Part II. Containing, a plain and comprehensive grammar….* Hartford: Printed by Hudson & Goodwin.

Webster, Noah (1790). *Rudiments of English grammar; being an introduction to the second part of the grammatical institute.* Hartford: Printed by Elisha Babcock.

Webster, Noah (1807). *A Philosophical and Practical Grammar of the English Language.* New Haven: Brisbon and Brannan.

Weinreich, Uriel, William Labov and Marvin I. Herzog (1968). "Empirical Foundations for a Theory of Language Change." Ed. Winfred P. Lehmann, *Directions for historical linguistics*: *A symposium*. Austin: University of Texas Press. 95–188.

Widlitzki, Bianca and Magnus Huber (2016). “Taboo language and swearing in eighteenth and nineteenth century English: A diachronic study based on the Old Bailey Corpus.” Eds. María J. López-Couso, Belén Méndez-Naya, Paloma Núñez-Pertejo and Ignacio M. Palacios-Martínez, *Corpus linguistics on the move. Exploring and understanding English through corpora*. Leiden: Amsterdam & New York: Brill/Rodopi. 313–336.

Wilson, John and Alison Henry (1998). “Parameter Setting within a Socially Realistic Linguistics.” *Language in Society* 27 (1): 1–21.

Withers, Philip (1790). *Aristarchus, or the principles of composition…,* [1788]. 2nd edn. London: Printed by J. Moore, Drury Lane.

Wolf, Göran (2011). *Englische Grammatikschreibung 1600-1900*: *Der Wandel einer Diskurstradition.* Frankfurt am Main: Lang.

Wolfson, Nessa (1979). “The Conversational Historical Present Alternation.” *Language* 55 (1): 168–182.

Wolfson, Nessa (1982). *CHP: The Conversational Historical Present in American English Narrative.* Dordrecht: Foris.

Yáñez-Bouza, Nuria and María Rodríguez-Gil (2010). *Eighteenth-Century English Grammars Database (ECEG).* <http://www.alc.manchester.ac.uk/subjects/lel/research/completedprojects/c18englishgrammars/>.

Appendix

A. Additional tables

A-1: Absolute frequencies underlying Figure 27: Observed proportions of HAVE**, by verb and period, OBC (only for verbs that occur at least five times/period), N = 9,916**

period	verb	BE	HAVE
1720-1769	COME	183	95
	GET	219	113
	GO	1,679	201
	PASS	0	34
	RETURN	2	3
	RUN	54	40
	TURN	14	6
1770-1819	COME	134	210
	GET	48	259
	GO	1,626	273
	PASS	0	68
	RETURN	8	10
	RUN	9	63
	TURN	6	28
1820-1869	COME	43	487
	GET	4	331
	GO	1,222	541
	PASS	0	79
	RETURN	0	16
	RUN	2	95
	TURN	11	28
1870-1913	COME	9	451
	GET	9	159
	GO	231	694
	PASS	0	53
	RETURN	1	13
	RUN	1	37
	TURN	1	13

A-2: Absolute frequencies underlying Figure 29: Observed proportions of HAVE, by decade, OBC (N = 9,982)

decade	BE	HAVE	% of HAVE
1720s	81	9	10.0%
1730s	567	127	18.3%
1740s	453	140	23.6%
1750s	534	102	16.0%
1760s	516	119	18.7%
1770s	414	177	29.9%
1780s	364	162	30.8%
1790s	296	174	37.0%
1800s	366	194	34.6%
1810s	391	218	35.8%
1820s	405	329	44.8%
1830s	309	322	51.0%
1840s	275	299	52.1%
1850s	186	314	62.8%
1860s	107	338	76.0%
1870s	82	291	78.0%
1880s	57	306	84.3%
1890s	61	345	85.0%
1900s	42	337	88.9%
1910s	10	163	94.2%

A-3: Absolute frequencies underlying Figure 41: Percentage of *I says* in the OBC, by decade

decade	*I said*	*I says*	*% I says*
1720s	5	79	94.1%
1730s	165	481	74.5%
1740s	617	129	17.3%
1750s	722	3	0.4%
1760s	882	3	0.3%
1770s	388	1	0.3%
1780s	392	184	31.9%
1790s	326	211	39.3%
1800s	364	40	9.9%
1810s	367	0	0.0%
1820s	516	0	0.0%
1830s	663	1	0.2%
1840s	827	0	0.0%
1850s	787	0	0.0%
1860s	1,112	8	0.7%
1870s	1,040	5	0.5%
1880s	1,443	3	0.2%
1890s	1,096	0	0.0%
1900s	1,022	2	0.2%
1910s	507	0	0.0%

A-4: Absolute frequencies underlying Figure 42: Discourse introducers with *says* p100tw in the OBC between 1720 and 1809: *I says* - *he says* - other (including other pronouns and NPs + *says*) (N = 3,716)

decade	*I says*		*he says*		other	
	abs.	p100tw	abs.	p100tw	abs.	p100tw
1720s	97	134.35	79	108.92	96	132.97
1730s	551	77.06	574	80.27	538	75.24
1740s	126	19.98	135	21.40	158	25.05
1750s	3	0.45	4	0.60	5	0.75
1760s	2	0.30	4	0.59	0	0.00
1770s	1	0.16	12	1.88	4	0.63
1780s	181	31.09	285	48.95	115	19.75
1790s	211	37.03	261	45.80	89	15.62
1800s	40	6.67	58	9.67	25	4.17

B. Scribes, printers and publishers of the *Proceedings of the Old Bailey*

B-1: Scribes and printers, based on Huber (2007: 3.3.2) and OBC annotation (* = includes *Proceedings* where there is doubt as to the identities of scribes or printers)

from	to	scribe	printer
16781211			G. Hills
16920406			Thomas Braddyl
17261207			J. Read
17381206	17401015		T. Cooper
17410405	17411014		J. Roberts
17411204	17420603		T. Payne
17420714	17430114		T. Cooper
17430223	17451016		M. Cooper
17451204	17460117		C. Nutt
17471209	17481207		M. Cooper
17490113	17551022	Thomas Gurney	M. Cooper
17551204	17571026	Thomas Gurney	J. Robinson
17571207	17591024	Thomas Gurney	M. Cooper
17591205	17601022	Thomas Gurney	G. Kearsley
17601204	17611021	Thomas Gurney	J. Scott
17730707	17750712	Joseph Gurney	
17751206	17771015	Joseph Gurney	William Richardson
17771203	17811017	Joseph Gurney	
17811205	17820410	William Blanchard	
17820515	17820703	Joseph Gurney	
17820911	17921031	E. Hodgson	
17921215	17951028	Manoah Sibly	Henry Fenwick
17951202	17970215	Marsom & Ramsey	W. Wilson
17970426	18011028	William Ramsey	W. Wilson*
18011202	18051030	Ramsey & Blanchard	W. Wilson*
18051204	18150510	Job Sibly	R. Butters
18150621	18160918	J.A. Dowling	R. Butters
18161204	18280110	Henry Buckler	T. Booth
18280221	18300415	Henry Buckler	Henry Stokes
18300527	18310512	Henry Buckler	Henry Stokes & George Titterton
18310630	18330411	Henry Buckler	George Titterton
18330516	18331128	Henry Buckler	William Johnston
18340102	18420704	Henry Buckler	William Tyler
18420822	18471025	Henry Buckler	Tyler & Reed
18471122	18490611	James Drover Barnett, Alexander Buckler	Tyler & Reed
18490702	18570406	James Drover Barnett, Alexander Buckler*	William Tyler*
18570511	19050109	James Drover Barnett, Alexander Buckler*	
19050206	19060205	Alfred Fitzgerald Dalton	
19060305	19130107	George Walpole	The Argus Printing Company

B-2: Publishers of the *Proceedings*, based on OBC annotation (* = includes *Proceedings* where information on publisher is not recoverable / doubtful)

from	to	publisher
17291203	17311013	T. Payne
17311208	17321011	J. Roberts
17321206	17341016	J. Wilford
17341204	17381011	J. Roberts
17381206	17401015	T. Cooper*
17401204	17411014	J. Roberts
17411204	17420603	T. Payne*
17420714	17430114	T. Cooper*
17430223	17451016	M. Cooper*
17451204	17461015	C. Nutt*
17461205	17471014	J. Hinton
17471209	17480115	M. Cooper*
17480116		M. Cooper, C. Davis
17480224	17551022	M. Cooper*
17551204	17571026	J. Robinson*
17571207	17591024	M. Cooper*
17591205	17601022	G. Kearsley*
17601204	17611021	J. Scott*
17611209	17631207	John Ryall*
17640113	17640222	E. Dilly*
17640502	17651016	
17651211	17681019	J. Wilkie*
17681207	17731020	
17731208	17741019	J. Williams*
17741207	17751018	
17751206	17761016	John Glyn
17761204	17771015	
17771203	17811017	Joseph Gurney
17811205	17820410	William Blanchard
17820515	17820703	Joseph Gurney
17820911	17921031	E. Hodgson
17921215	17951028	Henry Fenwick
17951202	18051030	W. Wilson*
18051204	18161030	R. Butters
18161204	18210606	T. Booth
18210718	18270913	T. Keys
18271025	18561027	G. Hebert*
18561124	18701024	Butterworths*
18701121	19060205	Stevens & Sons
19060305	19130401	George Walpole