CHAPTER 1

Introduction

1.1 Aims

The three aims of this book, which are accounted for in Sections 1.1.1–1.1.3, primarily concern the syntax of Late Modern English (henceforth "LModE"), that is, the English used between *c.*1700 and *c.*1900. However, the topics discussed in the book also have wider applications that are relevant to the history of English and historical linguistics in general. The aims can be roughly characterized as theoretical, methodological, and empirical, respectively, though as I hope to demonstrate, the three are intertwined. From a theoretical perspective, I aim at reconciling the oft-cited view that language change in LModE is more limited than in other periods with what social-network theory would predict by focussing on the individual speaker's idiolect as the locus of language and language change. I then demonstrate how this perspective is compatible with historical corpus linguistics as a methodology. Finally, I aim to empirically investigate two types of change – colloquialization and densification – in nineteenth-century English through corpus-linguistic case studies of four features: *not*-contraction, co-ordination by *and*, nominal premodifiers in noun phrases, and participle clauses as postmodifiers in noun phrases.

1.1.1 *Late Modern English and the Stability Paradox*

The first aim of the present study is to examine a claim that has frequently been made in scholarship on LModE, namely that the structure of LModE features little change over time when compared with previous periods. Romaine (1998b: 7), for instance, comments on the "structural stability" of LModE. This stability is typically taken to concern phonology and grammar, while the vocabulary of LModE – which is also less connected to the structure of the language – is considered to be more open to change

(Romaine 1998b: 1–2). In addition, it is sometimes argued that changes in phonology or grammar which did take place in LModE were of different types compared with those attested for previous periods, especially from the late eighteenth century on. Strang (1970: 78–9) argues that phonological change after *c.*1770 has been characterized by "the complex analogical relationship between different parts of the language" rather than by change to the system itself, and Denison (1998: 93) suggests that, after 1776, syntactic change has typically concerned statistical preferences for particular features in different styles rather than "categorical losses or innovations". I will argue that such statements are valid when considered from the perspective of the communal language that we refer to as LModE, but less applicable to the individual idiolects that made up that communal language.

As I demonstrate in Chapter 2, the claim that LModE features little change in its structure over time is seemingly at odds with what we know about the interaction between social networks and language use. The widespread technological and sociocultural transformations that Late Modern English society underwent are likely to have created a plethora of weak network ties at the expense of pre-existing strong ones. Such weak ties are in turn assumed to facilitate language change (see, for instance, Milroy and Milroy 1985). It seems that we need to revise at least one of three assumptions, since we cannot simultaneously argue that the LModE period featured weakened social-network ties, that weakened ties facilitate language change, and that little structural change took place in LModE. Romaine (1998b: 7) rightly points out that the appearance of stability in LModE "challenges any simple-minded view of the relationship between social change and language change". I refer to this conundrum as *the stability paradox*.

The suggestion that language change in LModE differed from that observed in previous periods in the history of English may seem to provide a way out of this difficulty: perhaps plenty of language change occurred, but without affecting the overall structure of LModE. However, as noted by Beal (2004: xii, 125), this suggestion constitutes a potential challenge to the uniformitarian principle, that is, the assumption that "the linguistic processes taking place around us are the same as those that have operated to produce the historical record" (Labov 1972: 101) and "that the general properties of language and of processes of change in language have been the same throughout human history and prehistory" (Matthews 2014, s.v. *uniformitarian principle*). This principle – essentially the extension of uniformitarianism from the sciences (originally geology) to linguistics –

makes it possible to assume that the same types of language change that are operative in the present occurred at some point in the past, unless it can be shown that the conditions under which English was acquired and used have changed between then and now. (By extension, the same principle holds between two points in the past, such as Middle English and LModE.) Refuting it is thus a strong claim.

While there has been a drastic increase in scholarly interest in LModE during the last twenty-five years, our new knowledge of language change between 1700 and 1900 needs to be integrated not only into our conception of how the English language has changed through time, but also into more general theories of how and why languages can, may, and do change. Not linking these fields of scholarship has negative consequences of two types. Deductively, research on LModE will suffer from being less informed by linguistic theory than are other periods in the history of the language; inductively, findings of potential theoretical significance in LModE studies may not be fully recognized outside this specialized field because they are not explicitly connected to linguistic theory. We thus need to engage explicitly with these issues, and this book is an attempt to do so. My suggestion, which is discussed in Chapter 3, is that a possible solution to the seeming mismatch between predictions based on social-network theory and attested change in LModE is to regard the idiolect as the true locus of language change – and to regard change on the idiolectal level as the actual correlate of weak network ties.

1.1.2 *Methodological Challenges for Studying the Syntax of Late Modern English (and Other Historical Language-States)*

The second aim of the present study is to reconcile the idiolectal perspective on the study of language change with historical corpus linguistics, which is the methodology applied in the case studies of nineteenth-century syntax. Historical linguistics presents researchers with a methodological challenge: scholarship is heavily dependent on empirical data, but, as Labov (1972: 100) notes, researchers "have no control over their data". In addition, with the exception of historical statements that happen to have survived, historical linguists lack access to native-speaker evidence. Even though LModE is close in time to the present day, it has to be reconstructed mainly based on surviving texts. The best-known difficulty resulting from this limitation is the necessity to rely on what Labov (1972: 100) termed "bad data", that is, data that "may be fragmentary, corrupted, or many times removed from the actual productions of native speakers". The

main problems with bad data from a linguistic perspective are (i) that information may be lacking on who produced a given text and/or on their gender, socio-economic rank, and so on, and (ii) that the selection of idiolects we have access to is biased in favour of male, literate, and/or high-status speakers.

The limitations outlined above apply to a greater or lesser extent to virtually any historical-linguistic methodology. From the idiolectal perspective on language change outlined in Chapter 3, historical corpus linguistics faces an additional problem: corpora are typically compiled with the aim of sampling several idiolects in order to limit the influence of individual language users on the overall results. To resolve this seemingly paradoxical state of affairs, my discussion in Chapter 4 will emphasize the importance of considering, among other things, the representativity of the primary sources used and the granularity of the analysis, that is, the trade-off between describing linguistic behaviour at a high level of detail in one idiolect (or a small number of idiolects) and describing linguistic behaviour at a lower level of detail in a large number of idiolects.

In addition to these points, I will also discuss and problematize several other aspects of historical (corpus) linguistics, including the choice between variationist and non-variationist designs and the interpretation of numerical data. In the organization of the book, the methodological discussion forms a conceptual bridge between the theoretical discussion and the empirical case studies.

1.1.3 Studying Colloquialization and Densification

The third aim of this book is to examine the occurrence of two types of language change in nineteenth-century English: colloquialization and densification. These are of course not the only types of change that characterize LModE syntax, but colloquialization and densification are interesting processes to consider from the perspective on language adopted in this book. The linguistic features involved are rarely the result of very recent categorical innovations; instead, they can be assumed already to have been available in virtually all idiolects, but their frequency in particular (sets of) genres increases in response to changes in genre norms. Those norms are in turn dependent on, among other things, the purposes and characteristics that writers intended for texts belonging to these genres to have. (Should they be accessible to a popular audience or mainly to specialists? Was their main purpose to entertain readers or to convey information efficiently? etc.) Within the constraints set by the genre

norms, language users have a great deal of freedom regarding whether and how often they wish to use a given feature, which means that idiolectal variation can be expected to occur in the data; as will become apparent especially in Chapters 7 and 8, the intended target audience – in these cases, of different newspapers – is also a relevant factor.

Several aspects of colloquialization and densification have been studied before, but we still need detailed case studies of individual linguistic features in LModE to understand the origin of these two types of change, both of which characterized twentieth-century English. In addition, a large number of genres need to be examined, as one consequence of both types of change is increased genre diversity in writing: some but not all written genres typically take part in colloquialization and densification. This type of study can contribute to our understanding of the interplay between idiolects, genres, society, and the communal language. Each type of change will be investigated mainly by means of two case studies.

Colloquialization, first discussed (to my knowledge) in Siemund (1995) and Mair and Hundt (1995), consists in a stylistic change "away from a written norm which is elaborated to maximal distance from speech and towards a written norm that is closer to spoken usage" and "away from a written norm which cultivates formality towards a norm which is tolerant of informality and even allows for anti-formality as a rhetorical strategy" (Mair 2006b: 187). It has been well documented for twentieth-century English by, for instance, Hundt and Mair (1999); studies such as Smitterberg (2008) and Biber and Gray (2012) demonstrate that it is also in evidence during the preceding century. In the present study, I juxtapose two linguistic features that share an association with orality but differ greatly in their stylistic markedness. *Not*-contraction, for example *won't* as opposed to *will not*, was stigmatized during the LModE period, and still remains an informal feature today. It can thus be expected that LModE language users would be aware of the colloquial status of *not*-contractions and that they may avoid using them in more formal contexts owing to prescriptive pressure. The other feature examined is the use of *and* to co-ordinate linguistic units above the phrase level, for example main clauses, subordinate clauses, or predicates, as in *I [went to the theatre] and [saw a comedy]*, as opposed to units on or below the phrase level, for example noun phrases in *I bought [a suit] and [three shirts]*. While the proportions of these types of unit have been shown to differ greatly between speech and writing (e.g. Biber et al. 1999: §2.4.7.3), with phrasal co-ordination being characteristic of non-speech-related writing and vice versa, there is no

particular stigma attached to the use of *and* to link super-phrasal units.[1] Language users who were affected by colloquialization as a change from below may thus increasingly have used *and* to co-ordinate clausal and clause-like material in what Hundt and Mair dub "agile" genres, such as newspaper writing, even though they would have avoided *not*-contractions in the same texts. The juxtaposition of these two features thus creates a fruitful empirical basis for discussing colloquialization during the nineteenth century.

Densification is a tendency to use less linguistic material to express a given semantic content over time (Leech et al. 2009: 249). It has mostly been studied as a condensation of information in the noun phrase. As noted by Biber and Gray (2012: 316), although in many respects this can be seen as an "anti-colloquial trend" (Leech et al. 2009: 210), colloquialization and densification are not mutually exclusive; they notably co-occur in newspaper language, for instance. In the present study, I consider two features that contribute to densification of content: nouns as premodifiers in noun phrases, as in *a book room*, and participle clauses as postmodifiers in noun phrases, as in *the lady crossing the street*. Both features take up less space than their main competitors do: depending on the noun + noun combination, noun premodifiers may be in variation with, for instance, prepositional phrases (e.g. *a room for books*); the most obvious alternative to a participle clause is a finite relative clause (e.g. *the lady who was crossing the street*).

Both colloquialization and densification potentially raise important questions about methodology, which link the second and third aims of the present study. First, a great deal of valuable research has focussed on overall trends in large amounts of primary material where we can largely disregard the influence of individual users on the overall results. However, an alternative perspective, where smaller amounts of material are examined and differences due to individual people, newspapers, and so on are identified and analysed, is also of considerable value, especially if advanced and conservative users can tell us something about the propagation of changes. Secondly, several analyses in the book necessitate a choice between the two most frequent ways of measuring the frequency of syntactic features in texts, namely a variationist approach, where the incidence of each variant is measured against the frequency of the other variant(s), and what, following Biber et al. (2016), I shall call

[1] The main exception in this regard concerns sentence-initial *and*, which was widely proscribed in the nineteenth century; as I show in Section 6.3.2, this use of *and* also displays different genre distributions compared with super-phrasal *and* in general.

a *text-linguistic* approach, in which raw frequencies are normalized in relation to a coefficient, typically a set number of words, to make them independent of text length. The relative merits of these two perspectives are discussed in Section 4.4 and examined empirically in three of the case studies.

1.2 Limitations in Scope

As I suggest in Chapters 2 and 3, connections between societal and linguistic developments cannot be overlooked if we wish to answer the question of why some periods in the history of a language appear to feature less change – or different kinds of change – than others. This connection between linguistic and social developments necessitated focussing on one particular LModE society for which we have ample linguistic as well as extralinguistic data; England seemed a promising choice. The discussion of linguistic and societal change, and the case studies, will thus centre on English society between 1700 and 1900, to the exclusion of other parts of the English-speaking world. LModE displayed an immense increase in both the number of varieties of English outside England and the number of speakers of those varieties. On the British Isles, English spread at the expense of Celtic languages; for instance, the majority of the native population of Ireland switched to English during the nineteenth century (Hickey 2010d: 267). Outside Britain, while the establishment of an English-speaking community in North America predates 1700, the North American population grew rapidly during the LModE period: Bailey (2010: 185) suggests a growth from *c.*210,000 to over 5.2 million between 1700 and 1800.[2] In addition, many of the regions where inner-circle and outer-circle varieties of English are used today were first settled by speakers of English during the eighteenth and nineteenth centuries (see Hickey 2004b). To some extent, the increased interest in these varieties can be viewed as a linguistic reflection of "trends in colonial and (especially) postcolonial studies" (Smith 2015: 197). As World Englishes became an integral part of English linguistics, the origin and development of extra-territorial varieties also came to be regarded as an important area of historical research. A few examples of recent explorations of different aspects of such varieties, including syntax, are Hickey (2004a), Fritz

[2] This growth includes migrants from non-Anglophone regions and enslaved people who did not speak English on arrival in America. Conversely, there was a massive decline in the population of Native Americans, many of whom did not speak English.

(2007), Dollinger (2008), and Reuter (2017). Although there are good reasons for restricting the scope of the present study to England, any LModE variety deserves equally full treatment.

The bulk of this book is concerned with syntactic change. The theoretical discussion in Chapters 2 and 3 also includes lexis and phonology, as considering data from different areas enriches the account of connections between language and society. In contrast, much of the methodological discussion in Chapter 4, as well as all case studies in Chapters 5–8, focusses on syntactic features.

The case studies all concern nineteenth-century English. Several macro-level studies of a large number of features, such as Biber and Gray (2012), have examined developments between centuries; the complementary aim of the present study was to consider change within a narrower time frame by sampling several periods from the same century. It follows that it was necessary to limit the period covered by case studies to one century in order to make the amount of data manageable, and the 1800s was the natural choice for three reasons. First, colloquialization and/or densification have been described by scholars such as Hundt and Mair (1999), Mair (2006b), and Leech et al. (2009), as they took place in the 1900s. It is then a natural research question whether the time-depth of these changes can be extended to the preceding century. Secondly, Biber and Gray (2012: 326) associate the growing pressure towards popularization as well as "the rise of highly educated populations in specialized fields" with the "development of near-universal education", which is a feature of the nineteenth century. The nineteenth century is also the first century when all the standard *not*-contractions were used in print (Brainerd 1989 [1993]: 177), which makes the 1800s a suitable period for analyses. Thirdly, newspaper language is of special importance to the case studies, because it has been affected by both colloquialization and densification (Biber and Gray 2012: 316). This makes the nineteenth century particularly relevant to the discussion, as it was a pivotal phase in the development of the newspaper; Brown (1985: 1) even claims that "'[t]he news' as we understand it is a nineteenth-century creation". Taken together, these circumstances presented compelling arguments for focussing the empirical section of the volume on the 1800s.

Finally, the case studies are mainly based on corpora that may be thought fairly small by today's standards. Owing to a combination of factors, including the recency of the period and the number of literate speakers, the textual evidence for LModE is more copious than for any preceding period. More importantly, owing to recent advances in digitization, optical character recognition, and so on, the early twenty-first century

has witnessed an increased interest in – and availability of – very large corpora and other text collections,[3] which are based on electronically available material and provide previously unavailable windows onto linguistic variation in LModE. As Davies (2012: 163) notes, corpora like COHA (the Corpus of Historical American English) allow researchers to examine both features with lower frequencies and earlier states of change compared with smaller text collections. Results based on very large corpora can also shed light on more fine-grained linguistic environments. As I discuss in Chapter 4, however, there are several reasons why the availability of these corpus giants does not invalidate the use of smaller corpora such as those used in the present study, and the two main corpora used provided sufficient data for most analyses (see Section 4.5).

1.3 The Structure of This Book

The remainder of the book is structured as follows. In Chapters 2 and 3, I address the main theoretical aim of the study. Chapter 2 begins with an account of the role of social networks and language change; I then introduce the stability paradox by examining both the social structure of English society and the linguistic structure of English between 1700 and 1900. Chapter 3 is devoted to showing how an idiolect-centred perspective on language and language change can help to resolve the stability paradox. I also demonstrate why some of the most important changes that occurred during the LModE period have left surprisingly small traces in surviving texts, and discuss the role of linguistic and extralinguistic factors in language change.

Chapter 4 focusses on various aspects of methodology. I first discuss historical corpus linguistics and show how this methodology can be reconciled with an essentially idiolect-based view of language. I then address colloquialization and densification – and their connection to the case studies – in detail. The choice between a variationist and a text-linguistic approach to frequency is discussed in a separate subsection, since that choice has important consequences for several of the case studies. Finally, I describe a crucial aspect of any corpus-linguistic endeavour, namely, the corpora used for the case studies: a Corpus of Nineteenth-Century English (henceforth CONCE) and the Corpus of Nineteenth-Century Newspaper English (henceforth CNNE).

[3] Such collections include the extensive database of quotations in the *OED* (*Oxford English Dictionary*), which has been used as a makeshift corpus in some studies (see Hoffmann 2004 for critical discussion).

The four case studies are accounted for in separate chapters, the two chapters on colloquialization (Chapters 5 and 6) preceding the two that treat densification (Chapters 7 and 8). I begin by considering *not*-contraction in Chapter 5. A multifactorial analysis of the choice between contracted and uncontracted forms enables me to show which of the factors analysed have an independent effect on the proportion of contraction; I also consider a number of additional parameters, including *no*-negation and word order. The different units co-ordinated by *and* are the subject of Chapter 6, where the results are divided into two main parts: a quantitative analysis of the proportions of three main types of conjoins, followed by a smaller, partly qualitative study of sentence-initial *and*, a stigmatized feature of LModE writing.

Chapter 7 addresses one of the most conspicuous features of LModE and Present-Day English noun phrases compared with earlier periods: the prevalence of nouns in the premodifier slot. After looking at the frequency of common nouns (e.g. *a goods train*) and proper nouns (e.g. *the Bradford team*) as premodifiers, I focus on the semantics of the relation between common-noun premodifiers and their head nouns and the reference of the proper-noun premodifier. The last case study concerns participle clauses as noun-phrase postmodifiers (e.g. *presented in this study* in *the results presented in this study*). Both principles regarding the selection and classification of data and the choice between variationist and text-linguistic perspectives on the incidence of participle clauses are shown to influence the results. In addition, two features of the participle clause – whether the verb phrase includes a present or a past participle and whether the clause is restrictive or non-restrictive – are shown to divide up the data in four categories with partly different characteristics.

The book ends with a concluding discussion in Chapter 9. In this chapter, I return to the aims of the study and demonstrate what light the results shed on LModE and on the historical development of languages in general.