Introduction: Exploring nineteenth-century English – past and present perspectives

MERJA KYTÖ, MATS RYDÉN AND ERIK SMITTERBERG

1 Introduction

The structure and use of the English language has been studied, from both synchronic and diachronic perspectives, since the sixteenth century.[1] The result is that, today, English is probably the best researched language in the world. But the field is as unlimited as language itself, and therefore there will always be gaps in our knowledge of the historical development of English as well as of its time-bound, or synchronic, uses. In this respect, Late Modern English (1700–1950) has been given less scholarly attention than other periods in the history of English. This is particularly true of the nineteenth century and the first half of the twentieth century. The main reason why this period has been relatively ignored by historical linguists is presumably that at first sight it appears little different from Present-day English, resulting in the view that not much has happened in the language in the course of the last 200 years or so (for discussion, see Romaine 1998a: 7; Rydén 1979: 34; Rydén and Brorström 1987: 9). As Beal (2004: xi) points out, until the millennium the nineteenth century was also 'the last century' from a contemporary scholar's perspective. The recency of nineteenth-century developments may have added to the view that the language of this period was not an interesting topic for historical research, where the 'antiquity' of the English language has often been in focus.

However, knowledge of the immediate or recent past is often crucial for our understanding of the language of the present day. Thus it is important to connect research on earlier periods, including Late Modern English, and on Present-day English into a coherent account aiming at a synthesis of the historical development of the English language. Areas of research such as verb syntax and the enrichment of the lexicon would benefit from such a coherent treatment. Moreover, such research should cover both stability and change in language. Yet the nineteenth century 'remains largely an unexplored territory' in this context (Kytö, Rudanko and Smitterberg 2000: 85; cf. also Denison 1998: 92). However, there are a number of relevant monographs such as Arnaud (1973), Dekeyser (1975) and Smitterberg (2005). Romaine (1998b) has also contributed greatly to our

knowledge of nineteenth-century English, and Poutsma (1914–29) remains an important source of information. In addition, there are some recent overviews of nineteenth-century English, such as Bailey (1996) and Görlach (1998 and 1999). These studies invite rather than preclude further research, and above all, they indicate that there is a rising scholarly interest in nineteenth-century English.[2]

A few *contemporaneous* studies of nineteenth-century linguistic usage exist, such as Andersson (1892) (relative clauses), Ljunggren (1893–4) (*shall/will*), Palmgren (1896) (temporal clauses) and Western (1897) (*can/may/must*). Also worthy of mention in the context of nineteenth-century English studies are Koch's and Mätzner's grammars (first issued in the 1860s) and Sweet's *A New English Grammar* (published in 1891 and 1898), which is the first modern English grammar written by an Englishman.[3] The language of great nineteenth-century authors has also received some scholarly attention.[4] A study on metaphors like Stitt (1998), bridging the gap between linguistic and literary studies, should also be noted here.

In the form of ten specialized case studies, the present volume aims to provide an overview of some intriguing aspects of nineteenth-century English that will shed new light on the language of this period. For reasons discussed in section 2.2, the variety in focus in most of the case studies is the standard language used in nineteenth-century South-Eastern England. This Introduction addresses some central methodological issues involved, and outlines the case studies, with special reference to the ways in which they illuminate both stability and change in nineteenth-century English.

2 Corpus linguistics and nineteenth-century English

2.1 The corpus-based approach

Most of the studies included in the present volume are based on data drawn from electronic text or citation collections, which are now considered the mainstay of empirical linguistic research. For Present-day English, the compilation of the seminal Brown and LOB (Lancaster-Oslo/Bergen) corpora, representative of 1960s usage, provided an impetus for a huge upsurge in research. Their structure has since been paralleled by other corpus compilers in order to enable the study of short-term linguistic change (e.g. by comparisons with the 1990s LOB and Brown 'clones', Freiburg-LOB and Freiburg-Brown) as well as regional variation (e.g. the Kolhapur corpus and the Wellington Corpus of New Zealand English).[5] In the 1980s, the compilation of the Helsinki Corpus of English Texts, comprising c. 1.5 million words, was a landmark in the historical study of the English language, from the Old English period up until the early 1700s. Since then, there has been an ever-growing interest in the compilation of historical corpora of English. However, as yet, computerized corpora covering the 1800s (and the early 1900s) have not been many (valuable exceptions include the ARCHER Corpus (A Representative Corpus of Historical English Registers), which covers

but does not specifically focus on the nineteenth century, and the Corpus of Late Modern English Prose, which covers the period 1861–1919 and centres on letters and journals).[6] This lack of interest in compiling corpora devoted to nineteenth-century English is probably a reflection of, among other things, the above-mentioned deceptive similarity to Present-day English exhibited by the language of the 1800s.

As regards Present-day English, the possibility of collecting texts representative of the language as a whole, from a linguistic as well as an extralinguistic perspective, offers new possibilities for linguistic research. As implied above, for the period following 1960 we have corpora including a wide variety of spoken as well as written genres. This availability of relevant texts enables an approach to language variation that takes medium into account as an extralinguistic parameter. Thus the researcher need not rely on genres consisting of speech recorded in writing, intended to represent speech. The immediate descriptive and pedagogical applications of corpora covering learner English and/or a wide spectrum of native-speaker Present-day English also influence corpus compilation. Moreover, compared with earlier periods, it is easy to correlate present-day linguistic data with extralinguistic factors not only on the textual level (e.g. cross-genre studies), but also on the level of the language user. This is because the characteristics of individual language users with respect to parameters such as social network structure, socioeconomic status and education are comparatively easy to ascertain and can be coded for as part of the compilation process. Most such user-related variables are more difficult to code for within a diachronic framework, gender being the main exception.[7]

2.2 The nineteenth-century perspective

In addition to the need for analysing nineteenth-century English as a link between Present-day and earlier periods of English, there are also reasons internal to the nineteenth century that justify the study of this period. Some of them are related to the comparatively rich and varied textual material available to us from this century. Owing to the spread of literacy during the 1800s, we have access to written texts produced by a greater proportion of all language users than is the case for any preceding period (however, as Tony Fairman's contribution to the present volume shows, the concept of literacy itself needs to be addressed in more detail). This is especially true as regards female language users: the nineteenth century thus offers promising possibilities of investigating the gender variable as a factor in language change (see below). In addition, research on the Brown, LOB, Frown and FLOB corpora has shown that as short a time span as thirty years may be sufficient to observe changes in linguistic usage, provided that the researcher has access to a sufficiently large corpus. The possibility of correlating short-term linguistic change in nineteenth-century English with the many important sociopolitical developments that took place during this period further adds to the potential of this approach to the study of linguistic variation.

The nineteenth century is also characterized by genres becoming more and more diverse in their linguistic make-up; for instance, the language of formal expository genres like academic writing and that of informal non-expository genres such as private letters diverge increasingly (see Biber and Finegan 1997: 272–3). In addition, the nineteenth century was crucial to the development of some genres that were to become (or remain) highly influential in the century that followed. Not only private letters, but also newspaper language, the novel and scientific discourse belong to this group of genres. These characteristics make nineteenth-century English a vital period for researchers interested in genre and cross-genre studies from a synchronic as well as a diachronic perspective.

Finally, the 1800s constitute a formative period in the development of many extraterritorial varieties of English. The nineteenth century is important with regard to the divergence of American and Canadian English from British English, and probably even more important as regards the development of Southern Hemisphere varieties of English. A full description of nineteenth-century English thus requires a broad regional scope: results valid for one regional variety of nineteenth-century English cannot safely be claimed to hold for the English language as a whole.

However, focusing on one variety may provide scholars with a useful starting-point for comparisons. Accordingly, the majority of the case studies included in the present volume are based on Standard English English, the most extensively researched variety across the centuries so far. From a diachronic perspective, they thus further our understanding of the development of Standard Englishes. At the same time, in synchronic terms, they help to establish a background against which nineteenth-century extraterritorial and non-standard varieties can be contrasted. A study such as Tony Fairman's (this volume) points to the potential of such a comparative perspective.

2.3 The CONCE project and the present volume

Given the wide range of research possibilities and the shortage of available corpora, any one corpus project must be selective in terms of attempting to capture the spectra of variation existing in nineteenth-century English. The present volume is, for the most part, a result of one such corpus project, launched at the Departments of English at Uppsala University and the University of Tampere in the mid-1990s. The aim of the project was to compile CONCE (A Corpus of Nineteenth-century English), a one-million-word corpus focusing on English English, and to produce research based on this new source of linguistic data.

Regardless of which period in the history of English researchers focus on, they are likely to take an interest in language variation and change. In investigating language change, scholars may carry out a synchronic study of a past stage of the language, such as nineteenth-century English, for comparison forwards and/or backwards in time. In this perspective, they may use CONCE as a synchronic whole and compare the results with those attested for, say, Early Modern English

Table 0.1. *Word counts for period samples in CONCE and for the whole corpus, excluding the words within reference codes and text-level codes.*

Period	Total
1 (1800–1830)	346,176
2 (1850–1870)	341,842
3 (1870–1900)	298,796
Total	986,814

or Present-day English. In the present book, this research strategy is used by, for instance, Juhani Rudanko, in his study of the pattern *in -ing* in the nineteenth and twentieth centuries.

However, scholars may also chart the development of a linguistic feature within a single historical stage of the English language. As mentioned above, recent studies based on the LOB, FLOB, Brown and Frown corpora are indicative of an increasing interest in such short-term linguistic change. Reflecting this research interest, the texts in CONCE have been stratified into three subperiods, covering, broadly speaking, the beginning, middle and end of the nineteenth century, viz. 1800–30 (period 1), 1850–70 (period 2), and 1870–1900 (period 3). This division makes it possible to study short-term developments across the 1800s.

The texts in CONCE have been coded using text-level and reference codes based on those applied to the Helsinki Corpus (see Kytö 1996b), making it possible to exclude, for instance, foreign language and headings from the counts. However, the system used for CONCE is slightly more rigorous, enabling the exclusion of passages such as stage directions in plays (see the Appendix for a full list of text-level codes). Table 0.1 presents word counts per period for the CONCE corpus.[8]

The division of the texts in CONCE into periods of several decades rather than single years, as with corpora such as LOB and FLOB, was a necessary compromise between the interests of (a) obtaining a sufficiently narrow periodization for cross-period comparisons to be reliable, and (b) including only texts that reflect authentic nineteenth-century English. In addition, important extralinguistic developments in England during the nineteenth century were taken into account. For instance, period 1 predates most of the political reforms of nineteenth-century England (e.g. the Reform Bills), while period 3 follows many of them (see Kytö, Rudanko and Smitterberg 2000: 87). Most studies in the present volume make use of this periodization in attempts to reveal diachronic variation within nineteenth-century English; for instance, Peter Grund and Terry Walker's study of the subjunctive in adverbial clauses traces the development of this verb form in relation to indicatives and modal auxiliaries across the century.

Table 0.2. *Description of the genres in CONCE.*

Genre	Characteristics
Debates	Recorded debates from the Houses of Parliament
Trials	Trial proceedings (in dialogue format)
Drama	Prose comedies including farces
Fiction	Novels
Letters	Personal letters (between relatives or close friends)
History	Historical monographs
Science	Monographs pertaining to the natural or social sciences

However, merely looking at texts from different historical periods of English may not be enough to identify the locus of the change under scrutiny, as linguistic change is frequently mediated through other extralinguistic parameters as well as time. Considering such parameters becomes even more important when relevant extralinguistic developments have taken place within and/or between the period(s) covered by the investigation. In the present volume, two extralinguistic parameters receive special attention: genre and gender.

Multi-feature/multi-dimensional analyses of both Present-day English and historical stages of the language, such as Biber (1988), Biber and Finegan (1997) and Geisler (2002, 2003), have shown that the frequency of a large number of linguistic features co-varies in texts, so that a given genre is characterized by different co-occurrence patterns along dimensions of linguistic variation. These patterns may also change across time, and Biber and Finegan's (1997) study indicates that the nineteenth century is of central importance in displaying increasing *genre diversity* in terms of linguistic make-up. The divisions between, on the one hand, oral, popular and/or non-expository genres, and, on the other hand, literate, specialized and/or expository genres are of particular relevance in this respect. In some cases, differences in the distribution of linguistic features across the genre parameter can also be used as a cross-section of linguistic change, with advanced genres representing a later stage in the development. However, accounting for this diversity requires that the researcher sample a range of genres. This requirement was one of the criteria used in the compilation of CONCE.

The original aim of the CONCE project was to provide comparative nineteenth-century follow-up material to the Helsinki Corpus of English Texts. Consequently, several genres are present in both corpora in order to increase comparability. In particular, it was deemed important to sample both speech-related and non-speech-related genres, and both formal and informal written genres, in order to enable research on how the use of nineteenth-century English varied according to the parameters of medium and formality. A brief description of the seven genres included in CONCE is given in table 0.2 (from Kytö, Rudanko and Smitterberg 2000: 88).

Table 0.3. *Word counts for period, genre and period/genre subsamples in CONCE and for the whole corpus, excluding the words within reference codes and text-level codes.*

Period	Debates	Trials	Drama	Fiction	Letters	History	Science	Total
1	19,908	62,360	31,311	42,032	121,624	30,904	38,037	346,176
2	19,385	60,570	29,543	39,045	131,116	30,504	31,679	341,842
3	19,947	67,588	29,090	30,113	90,891	30,564	30,603	298,796
Total	59,240	190,518	89,944	111,190	343,631	91,972	100,319	986,814

In addition, the fact that the texts in CONCE have been stratified with respect to both time and genre makes it possible to combine these two parameters in analyses of how linguistic change is reflected across the genre parameter. As implied above, for some changes this will result in a diversified picture with some genres being more advanced than others for each particular change. Table 0.3 (from Kytö, Rudanko and Smitterberg 2000: 89) presents word counts by period and genre for the CONCE corpus (for the full list of source texts, see the Appendix).

Reflecting the make-up of CONCE as well as the importance of the genre parameter, many of the contributions to this book focus on cross-genre variation. Among others, Christine Johansson considers three of the genres in CONCE in an analysis of the use of relative clauses in nineteenth-century English.

As regards gender, the rich textual material available from the nineteenth century makes it possible to investigate the interaction of several factors that, according to previous research, increase differences between female and male usage. Labov (2001: 292–3) claims that, on the whole, women will conform more than men to prestige norms if these norms are specified overtly; conversely, in processes of change from below, which take place below the level of normative consciousness, women tend to be leaders in linguistic change. There are several ways in which nineteenth-century English is an excellent testing ground for comparisons of these different influences on female and male usage. First, as mentioned above, the percentage of female literates increases dramatically during the 1800s (see Altick 1957: 171). Consequently, texts produced by women are more readily available in the nineteenth century than previously. Secondly, nineteenth-century attitudes to language variation embodied a largely prescriptive attitude on the part of grammarians. This attitude resulted in a number of grammars in which some lexical and morphosyntactic variants were promoted at the expense of others (see e.g. Dekeyser 1975, who compares precept and usage as regards number and case relations in nineteenth-century English, and Denison 1998: 150–8, who studies the emergence and diffusion of the progressive passive). The 1800s thus afford more data concerning

Table 0.4. *Word counts for the letters by female and male writers in CONCE, excluding the words within reference codes and text-level codes.*

Period	Female	Male
1	69,271	52,353
2	62,340	68,776
3	50,154	40,737
Total	181,765	161,866

both usage and precept than most other periods in the history of English (see section 2.2).

Reflecting the importance of gender aspects in nineteenth-century English, CONCE was compiled to enable a gender perspective on language variation and change. In an effort to include both women's and men's voices, the Letters genre has been stratified in order to include the same number of texts by female and male letter-writers. The necessity of reducing idiolectal influence on the overall figures meant that more informants were sampled for the Letters genre than for the other genres: each period includes five texts by female letter-writers and five by male letter-writers.[9] The Letters genre may be regarded as especially suitable for studying the interaction of the gender parameter, change from above and below and linguistic precept. It constitutes a written category that is influenced by spoken and/or colloquial norms; moreover, in general, private letters are not normally intended for publication. Considering these production circumstances, the texts in Letters can be expected to contain both language that has been influenced by the norms promoted in grammars, and less self-monitored language. Word counts by period and gender in Letters are given in table 0.4 (from Kytö, Rudanko and Smitterberg 2000: 90).

Strictly, the letter-writers have in fact been coded according to their biological sex rather than their socioculturally established gender identity. However, given that the differences in language use between women and men attested in the CONCE data are likely to be due to gender rather than sex, and in the interests of simplicity and consistency, we will use the term 'gender' to refer to the parameter. Many of the studies contained in this volume investigate linguistic variation with gender by comparing female and male usage; for instance, Ingegerd Bäcklund looks at how terms for women and men were modified linguistically in nineteenth-century English.

Needless to say, the parameters of genre and gender can be studied simultaneously, and several studies in the present volume combine the two. This approach has great potential, as a cross-genre comparison may reveal stylistic grading in the distribution of a linguistic feature, a finding that can then serve as the background

for a study of differentiation between women's and men's language. Merja Kytö and Suzanne Romaine use this perspective in an analysis of adjective comparison in CONCE.

However, a volume such as this, which aims to provide an overview of nineteenth-century English from several angles, cannot rely solely on extant corpora. There are both linguistic and extralinguistic reasons for extending the scope of the material studied beyond that currently covered by corpora. To begin with, the study of low-frequency features may reveal results that are of great theoretical and methodological significance; but a historical corpus like CONCE does not, on its own, provide sufficient material for all such investigations. In addition, in spite of the enlargement of the franchise and the spread of literacy during the 1800s, the texts which have been sampled in electronic corpora by and large reflect the language used by the upper echelons of nineteenth-century society. In order to provide a fuller picture, it is therefore necessary to sample texts outside those available in computerized corpora. Christian Mair's and Tony Fairman's contributions reflect this need to go beyond corpora when approaching particular research questions. Basing his research on the *OED* on CD-ROM, with its vast quotation database, Mair investigates verb complementation after *remember*. Fairman has compiled his own material, which includes pauper letters written chiefly between 1800 and 1834, in order to investigate the written language of these, often barely literate, letter-writers and compare some aspects of their production with those attested in letters by more fully schooled and wealthier people.

3 Empirical evidence of nineteenth-century English: stability and change

The studies included in this volume illustrate nineteenth-century English on several levels. As discussed in section 1, nineteenth-century English is of particular interest because it displays what may be termed a deceptive similarity to present-day usage in many respects. Some features of the English language have remained stable in the past 200 years; others, however, have developed over time. Reflecting this tension between stability and change, some studies included in the present work did not reveal evidence of language change in progress, pointing rather to stability and linguistic continuity. Other studies, in contrast, unearthed new evidence of differences across time.

In diachronic studies, the focus is often on linguistic change. However, linguistic stability is also an essential object of study, as the possibility of tracing conditioning factors, both linguistic and extralinguistic, allows comparison of the situations that appear to encourage change with those which seem to promote stability (Raumolin-Brunberg 2002: 102; see also Rydén 1979: 19). Linguistic stability is not a concept exclusive to nineteenth-century English: as Raumolin-Brunberg (2002) has shown, linguistic features may exhibit stability for centuries. However, the 1800s offer linguists excellent opportunities of relating their

results to extralinguistic parameters, such as sociopolitical events and stylistic judgements (as observed in style manuals, grammar books, etc.), that may promote either change or stability. Given its importance, the stability vs change parameter is taken into account in the overview (below) of the studies in this volume.

It is of course difficult to group studies as focusing on either linguistic stability or change: many diachronic investigations yield evidence of both, depending on factors such as the level of analysis chosen (e.g. subperiodization scheme, the range of genres and informants included, and the level of detail in linguistic analyses). Nevertheless, a tentative grouping of the studies can be established on this parameter, based on dominating trends in the data. Of the ten studies included in the volume, three present results that highlight stability rather than change, as against the five that emphasize change over stability. In two studies, the starting-point and/or the results do not justify a conclusive classification in this respect. These three groups of studies will be presented in that order below.

Drawing for data chiefly on the Science genre of CONCE, Larisa Oldireva Gustafsson shows that continuity rather than change characterizes the use of the passive in nineteenth-century scientific writing, although there is a great deal of variation among the authors represented. In fact, this variation is found to underlie some of the apparent changes in Gustafsson's data, such as the higher frequency of passives in simple sentences towards the end of the century (subordinate clauses constitute the most frequent locus of passives in all periods, however). Gustafsson shows that there are also considerable differences in usage between scientific texts and private letters written by the same person (Charles Darwin), as also found by Mark Kaunisto with regard to *that of/those of* constructions.

The overall frequency of the passive in the Science genre remains stable across the 1800s, despite the variation among authors mentioned above. Neither does the distribution of the passive across the parameters of tense, aspect and mood seem to change dramatically across the 1800s, small increases in the relative frequency of future passives, perfect passives and indicative passives notwithstanding: present-tense indicative passives that are unmarked for aspect dominate the distribution in all periods. There is shown to be a good deal of continuity with Present-day English regarding which main verbs are used most commonly in passives in scientific writing. The author concludes that the development of the passive as a characteristic feature of English scientific writing is most likely to have taken place before the nineteenth century. The results also imply that genre is a more important parameter than time regarding the use of the passive from 1800 onwards.

Christine Johansson analyses the use of relativizers in nineteenth-century English, and the distribution of *that* and *wh*-forms (*who*, *whose*, *whom* and *which*) in particular. She shows that at least in Trials, Science and Letters, *wh*-forms are used much more frequently than the relativizer *that*. In Present-day English, *that* has been gaining ground, but this development was still in