Register in historical linguistics

Merja Kytö
Uppsala University

Merja Kytö is Professor of English Language at Uppsala University. In this article, she provides a detailed accounting of the role of register in research on the historical development of language. Her substantial body of work has focused on both the historical development of specific registers, as well as how historical change has been mediated by register. Her research has encompassed a range of time periods (from Early Modern English to the 19th century) and registers (for example, depositions, Salem witchcraft records, and dialogues). Her many edited collections have brought historical linguists together into comprehensive and rigorous volumes, including the *Cambridge Handbook of English Historical Linguistics* (Kytö & Pahta 2016, Cambridge University Press), *English in Transition: Corpus-Based Studies in Linguistic Variation and Genre Styles* (Rissanen, Kytö, & Heikkonen 1997, De Gruyter), and *Developments in English: Expanding Electronic Evidence* (Taavitsainen, Kytö, Claridge, & Smith 2014, Cambridge University Press). She has been a key contributor to the development of principled historical corpora, such as the *Helsinki Corpus of English Texts* project, which represents a range of registers from Old and Middle English to Early Modern English. Merja Kytö has long been a leader in demonstrating how systematic attention to register can result in rich profiles of historical development, and in addressing the inherent challenges involved in utilizing historical documents for linguistic research.

Keywords: language change, historical corpora, genre, variationist, multi-dimensional analysis, sociolinguistics, grammaticalization, colloquialization, standardization

1. How is register conceptualized in historical linguistics?

The importance of registers and register-related notions such as genre and text type as powerful factors shaping language use and linguistic and stylistic change has been acknowledged widely in historical linguistics. Their relevance to historical corpora and historical corpus linguistics has also been recognized. Accord-

https://doi.org/10.1075/rs.18011.kyt
Register Studies 1:1 (2019), pp. 136–167. issn 2542-9477 | e-issn 2542-9485

ing to Wright (1994:101), the use of registers is of key importance to a historical corpus: "it is the crucial axis in the design of most existing text corpora for historical linguistic analysis". It is often on the basis of extralinguistic criteria such as textual functions, presentation conventions, and audience expectations that texts have been grouped under terms 'genre' or 'text type' (Claridge 2012:238; Kytö & Smitterberg 2015:332). On the other hand, the term 'text type' has been applied to classifications based on text-internal linguistic features by some scholars (Biber 1988:170; Taavitsainen 1997:187–190; 2001). 'Register', then, is rather a superordinate term that has been used loosely for classifications based on external criteria, or when distributions of linguistic features in texts are explored as a basis for register description. According to Biber and Conrad (2009:6), the essentials of register description comprise the situational context, the linguistic features, and the functional relationships between them; by situational context is meant here, e.g. communicative purpose, channel, participants and their relationships, and topics.

These notions have been approached from both language-theoretical and empirical perspectives, and a number of the strands of research can be traced back to literary studies. Most researchers are aware that in empirical studies the generalizability of their results largely depends on the register(s) or genres their data represent. Yet, research on the dynamics and diachronic development of registers is rather a neglected area in mainstream research on the history of English (for discussion, see Taavitsainen 2016:271).

A theoretical framework promoting the empirical study of language variation and change was Weinreich, Labov, & Herzog (1968), who argued for linguistic heterogeneity being the prerequisite for language change and the need for systematic study of changes in the patterns of variant forms across the history of a language. Within this framework, time and register are among the main extralinguistic factors potentially influencing trends of change. The advent of electronic corpora and corpus linguistic methodology greatly promoted the applicability of the variationist framework.

Over the past few decades, literary scholars have shown a keen interest in register and genre studies, among them, e.g., Fowler (1982), Bakhtin (1986/1953), and Todorov (1990). Research on literary and non-literary registers and genres has also offered a fertile ground for interaction between literary scholarship and linguistics. One of the key questions has been how texts can be grouped to form 'families' or 'classes'; are these categories rigid and clear-cut or are they fuzzy and flexible (for discussion, see Taavitsainen 2016:272–273)? For instance, Fowler (1982) advocated prototype approach to literary genre studies with fuzzy boundaries between the categories distinguished. The prototype approach applies well to the study of register or genre dynamics and mechanisms of change as, in addition to new or evolving features, there are usually also features that remain stable and ensure the

continuity of the register or genre in question. Another approach, especially in linguistics, has been to look into genre labels such as 'cookbook', 'private letter' or 'sermon' that reflect textual functions (see, e.g., Görlach 2004). Within this framework it becomes obvious that new genres can emerge in the course of history (e.g. news writing) while others may fall into oblivion (e.g. medieval verse romances). Or, a register may go through a change across the history owing to diversification of genre functions (e.g. legal writing and its sub-genres); this makes it necessary to observe the effects of possible internal register or genre change in analyses of material spanning across several centuries. Grouping genre labels under functional umbrella categories has also been introduced as a viable approach. For instance, Werlich (1982) distinguishes five basic categories: descriptive, narrative, expository, argumentative, and instructive. This classification was used as a basis for, e.g., defining the prototypical text categories in the *Helsinki Corpus*.

An important advance in historical register studies has been the introduction of multi-dimensional methodology which was first applied to present-day data in Biber (1988). In this synchronic study, occurrence patterns of linguistic features underlying dimensions of variation were interpreted to reflect different communicative functions. Such factor analyses have subsequently been applied to trace back trends of development across the historical varieties of English (see, e.g., Biber & Finegan 1989 and 1997 for Early and Late Modern English; Meurman-Solin 1993 for Older Scots; Taavitsainen 1997 for Early Modern English; Biber & Gray 2013 for Late Modern and Present-day English).

Further advances in modern language-theoretical or other analytical frameworks have also informed work on historical registers. For instance, within Systemic-Functional Linguistics, Halliday (1988) envisaged a trajectory of evolution for scientific writing, emphasizing the variability and continuous diachronic change in the development. In historical pragmatics, several central notions involving awareness of register issues can be traced back to analytical models developed for studies of modern language use or socio-cultural phenomena. Notions such as 'discourse community' (Swales 1990) and 'network of genres' (Swales 2004) were developed for studies within applied linguistics. Similarly, the notion of 'communities of practice' (CoPs), which has been applied to historical language data successfully, was originally developed for the purposes of anthropology and education studies (Lave & Wenger 1991; Wenger 1998). Within rhetorics of science, genres are considered important meaning-making vehicles of communication that echo the needs of authors and audiences (Bazerman & Paradis 1991; see Taavitsainen 2016: 278–279): changes in discourse communities are likely to lead to changes in register and genre conventions when authors wish to ensure that their texts communicate their thoughts in ways that are palatable to their audiences. Regarding mechanisms and processes of change,

grammaticalization and lexicalization phenomena have been among the well-researched areas where the register aspect has also been paid attention to.

Although the present discussion draws mostly on the work done on the English language, register, genre, and text type have inspired important historical-linguistic research on languages other than English as well. Regarding German, for instance, notable examples include von Polenz (1999), which provides a comprehensive account of recent research literature, Lange (2009), which focuses on texts and text types in the history of German, and Hauser, Roth, and Kleinberger (2014), which addresses current topics in diachronic text type linguistics, among them change in text patterns and text types. As for Romance languages, the term "register" is hardly ever used in diachronic studies. Instead, issues having to do with registers from the historical perspective are treated above all within diasystemic variation studies using terms such as 'genre', 'discourse tradition', or 'diaphasic variety'. Genre and discourse tradition are closely related notions within the diachronic framework in Romance linguistics, and a genre can be defined at many levels of abstraction; for a current survey, see Kabatek (2018). In addition to French, attention has also been paid to other Romance languages in historical genre studies, e.g., Spanish. In their study of modal constructions in older Spanish, Arroyo, Luis, and Schulte (2017) draw on a diachronic corpus of personal correspondence and adopting a diachronic variationist approach, trace back the influence of morphological, syntactic, semantic, and stylistic factors on the choice between the competing modal periphrases. Their choice of the material aimed at reflecting actual language use, personal correspondence, is indicative of the influential approach focusing on how to reconcile conceptual and medial orality and literacy presented by Koch and Oesterreicher (1985). The scales of communicative immediacy highlighted in this approach have inspired a good deal of research not only on Romance languages but also on German and, recently, on English (for examples, see below).

2. How does register relate to the research goals within historical linguistics?

The spectrum of research on historical linguistics is wide, extending from philological studies comprising, e.g. manuscript studies, paleography, text editing, layout, and paratextual aspects to linguistic analyses of data informed by various language-theoretical positions such as variationist sociolinguistics; historical sociolinguistics; historical pragmatics (and socio-pragmatics); historical phonology, morphology, syntax, semantics, and lexis; construction grammar; generative approaches; and multilingualism. Before the 1980s, empirical research was mainly

understood as the use of incidental examples from haphazardly collected texts. After the 1980s, propelled by the increasing interest in language variation and change (cf. Weinreich, Labov, & Herzog 1968) and in advances in corpus linguistic techniques, a good deal of research in English historical linguistics has been empirical in nature and aimed at accounting for patterns of linguistic variation and change with data drawn from extensive electronic corpora. Rissanen (2012: 198) goes as far as to say that "without the support and new impetus provided by corpora, evidence-based historical linguistics would have been close to the end of its life-span in these days of rapid-changing life and research, increasing competition on the academic career track and the methodological attraction offered to young scholars".

In the research carried out in the above vein, both quantitative and qualitative approaches are applied, and to an increasing extent, in combination. Registers and genres are often taken into account as potential factors influencing change, but they can also serve as the object of study *per se*. The research carried out so far has shown that rather than one unidirectional line of development across registers and genres from the Middle Ages onward, there have been diverse lines of development responding to changes in sociocultural structures, readership needs, and other external factors (Taavitsainen 2016: 271).

The register or genre factor has been treated in diachronic studies in two ways. Either several registers have been sampled for each period investigated, the overall expectation being that the entire language or language variety should be represented (Leech 2007: 135), or data collection has been limited to particular registers and their sub-registers (Biber & Gray 2011). Regarding the former approach, stratified multi-genre historical corpora suited for study include, e.g., the *Helsinki Corpus* and ARCHER, which comprise texts representing a many-sided range of registers extending from formal to informal language, speech-related contexts and written language, interactive and other modes, with a sub-periodization scheme justified by historical events or other principles (for the principles applied to, e.g., the Early Modern English section of the *Helsinki Corpus*, see Nevalainen & Raumolin-Brunberg 1989, 1993).

This approach has been challenged as it may not be very meaningful to attempt to represent historical language use in its entirety, considering, e.g., the gaps that there are in the survival of texts and the lack of audio-records of speech from the early periods. Instead, analyses focusing on particular registers have been proposed as a way of gaining insights into aspects of past and present language use. Although multi-genre corpora can certainly offer material for register-specific studies, textual representation may remain rather meager for individual registers. In this respect using register-specific corpora has proved a promising avenue. Among such resources are the families of corpora associated with the *Corpus of*

Early English Correspondence (CEEC) containing private and official letters and the *Corpus of Early English Medical Writing* (CEEM) providing access to a gamut of types of vernacular medical writing. Further specialized corpora include the *Corpus of English Dialogues 1560–1760* (CED) comprising genres representative of early speech-related language (trial records, witness depositions, drama, fiction, handbooks), and the *Zurich Newspaper English* (ZEN) corpus, which mirrors the diversification of newspapers from the late 17th to the 18th centuries. Even though these resources basically represent one main register, they are valuable in that they also provide access to several sub-registers. This enables cross-register comparisons which are important so as to account for change and put the results obtained into a wider perspective.

Studies based on several registers have offered convincing evidence of trends of development across the history of English. Among pioneering studies are Biber and Finegan (1989) and in the aftermath of that seminal article, follow-up studies by Biber and Finegan (1992, 1997), Atkinson (1992, 1996, 1999), Meurman-Solin (1993), Taavitsainen (1997) and Geisler (2002, 2003). For their 1989 study of the historical evolution of written genres of English, Biber and Finegan used a corpus comprising 115 texts (some 120,000 words) compiled by notable British and American authors. The period covered was divided into four sub-periods, i.e., the seventeenth, eighteenth and nineteenth centuries, and the modern period (from 1865 to ca. 1950). The three genres included in the corpus were essays, fiction and letters. By using a multi-dimensional approach based on factor analysis, Biber and Finegan empirically identified three central dimensions of linguistic variation revealing differences among literate and oral varieties. The authors interpreted the dimensions to convey values on the scales 'Informational versus Involved Production', 'Elaborated versus Situation-Dependent Reference', and 'Abstract versus Nonabstract Style'. There were differences in the rates and extents of change among the genres, but overall, the genres displayed a general drift towards more oral styles, i.e., more involved, less elaborated, and less abstract modes of presentation. The authors postulated that a number of functional and attitudinal factors could account for the development, among them, the rise of popular literacy and mass schooling, the demands of scientific and expository purposes, and conscious preferences by the writers of the texts. Significantly, this study paved the way for a new way of looking into register variation in historical linguistics. In their subsequent 1997 study, Biber and Finegan expanded the textual repertoire of their material and used a version of ARCHER (1650–1990) that comprised 1,037 texts, totaling 1.7 million words and representing seven written and three speech-based text categories. The material had been tagged for grammatical and functional features. This study demonstrated that increased linguistic diversity among written registers was typical of the Late Modern English period and that while,

indeed, personal styles of communication (journals/diaries and personal letters, prose fiction, news reportage) have increased in orality in their linguistic make-up since the nineteenth century, specialist expository registers (legal opinions, medical prose and scientific prose) have evolved in the opposite direction.

Further studies within the multi-dimensional framework comprise Atkinson's detailed 1992 and 1996 investigations of the evolution of two expository registers, i.e. medical and scientific research writing included in ARCHER. Examples of work on the *Helsinki Corpus* and its satellite corpora include Meurman-Solin (1993), who explored the evolution of prose genres in Older Scots, and Taavitsainen (1997), who looked into genre conventions used to convey personal affect in fiction and non-fiction in Early Modern English. Regarding the use of further corpora, in his multi-dimensional study of register variation based on *A Corpus of Nineteenth-century English* (CONCE), Geisler (2002) found that there were differences in how historical speech-related registers patterned together. Parliamentary debates align with the expository registers while trial proceedings appear non-expository; yet, both genres are speech-based records. In a further study on the same corpus, Geisler (2003) detected gender differences in nineteenth-century correspondence and demonstrated that while men's writing becomes more involved and narrative, women's writing becomes less involved and less narrative.

As for research on register-specific corpora, multivariate analysis combined with the study of frequency patterns was used by Nevalainen and Raumolin-Brunberg (2003), who investigated the role played by gender, social stratification, and regional variation in promoting linguistic change in Late Middle and Early Modern English correspondence. Drawing on evidence from CEEC, the authors looked into fourteen central grammatical developments that characterized the language in Tudor and Stuart times and reflected the incipient or on-going standardization phenomena. Among the features studied were the third-person endings on present-tense verbs, the loss of multiple negation, the development of *do*-periphrasis, possessive determiner *its*, and the prop-word *one*. The notion of register was introduced to the discussion of real-time language change and defined as "the relation between writer and addressee, the idea being that writers accommodate linguistically to their addressees" (Nevalainen & Raumolin-Brunberg 2003: 14). The findings for some features (e.g., the loss of the -n- in possessive determiners) indicated that the role played by register in a development can change over time, a result also obtained in earlier studies (e.g., Schendl 1997). However, it was either region or gender that drove the change overall, suggesting that with some exceptions, the degree of register variation in the personal correspondence data does not exceed the degree of gender or social variation (Nevalainen & Raumolin-Brunberg 2003: 199–200). In this respect, studies of historical written language vs. texts conveying glimpses at past spoken language are of particular interest. It should also be

mentioned that the social dimensions of register variation are still relatively rarely addressed in historical studies. Exceptions to this include the work done on the correspondence in the CEEC family corpora and in the CONCE corpus. The gender variable has also been investigated in the drama texts of ARCHER (e.g., Biber & Burges 2000), and in the same vein, social interaction in the play-texts and trial proceedings of the *Corpus of English Dialogues 1560–1760* has been among the targets of study (e.g., Culpeper & Kytö 2010), As for future prospects, the *Old Bailey Corpus* (OBC, version 2, 24 million words), which has been coded for speakers' gender, social class, and role in the courtroom interaction, can be expected to promote interest in the study of social dimensions of register variation.

Setting the multivariate techniques aside and focusing on qualitative analyses and frequency analyses of linguistic features (occurrence measured in proportions of variant forms or as incidence figures per a certain number of words), studies on multi-purpose and register-specific corpora have yielded new information on the role played by registers in the history of English. In their study of Early Modern English dialogues, combining methodology from the variationist approach, historical pragmatics, and historical sociopragmatics, Culpeper and Kytö (2010) investigated a number of individual linguistic variables characteristic of spoken face-to-face interaction (e.g., lexical repetitions, the conjunction *and*, grammatical features such as neuter possessive *its*, and the prop-word *one*) across the five speech-related genres included in the CED. They found that the play-texts emerged as the most speech-like of the genres even though none of the genres could be characterized as speech-like in every respect (2010: 401). Further, using the CED and a number of other corpora, Lutzky investigated the discourse markers *marry*, *well*, and *why*, adding letters, diaries, and sermons to the CED genres (Lutzky 2012). The study showed that the three discourse markers clustered in dialogic and fictional texts aimed at imitating speech and that they were less frequent in monologic texts and sub-registers based on records of authentic speech events. The author concluded that discourse markers were "prominent interactional features" and that "during the EModE period authors may have been aware of their particularly speech-like nature and hence included them specifically when constructing speech" (Lutzky 2012: 270).

While the above studies have all approached register variation from the perspective of a considerable or even large number of different linguistic features in texts, there is perhaps even a more extensive body of research focusing on a single linguistic variable (represented by several forms) and paying attention to qualitative differences or differences in the occurrence rates across the periods in multi-register or single-register corpora. Register is a central factor in such studies, serving as the frame for data collection and analyses. Studying individual features was possible even before the advent of electronic materials, and many such studies

still work as reference points to be taken into consideration when one approaches a topic from the perspective of modern corpus linguistic methodology. The register or genre variable is taken into consideration in these studies to a varying degree of systematicity. Among such studies is Ellegård (1953) on the history of *do*-periphrasis, Rydén (1966) on the use of relative constructions in Sir Thomas Elyot's and his contemporaries' writings in the sixteenth century, and Rissanen (1967) on the uses of *one* in Old and Early Middle English, a topic the author re-visited in his 1997 study based on the *Helsinki Corpus* (Rissanen 1997).

Among studies that draw on electronic historical corpora with focus on a linguistic variable and attention paid to register variation is Yáñez-Bouza (2015a) on the development of preposition placement in 1500–1900, with special focus on preposition stranding (P-stranding) (e.g., *The house which I live in*) vs. pied piping (P-piping) (e.g., *The house in which I live*). This study was based on the *Helsinki Corpus* and ARCHER and aimed at testing how well what was stated in late eighteenth-century normative works (precept) would tally with corpus evidence (usage data). Some of the register setups in the corpora were re-adjusted to harmonize the treatment, and in addition to the level of formality, the influence of spoken and written registers was investigated. P-stranding is mostly associated with informal and P-piping with formal registers (Yáñez-Bouza 2015a: 33). Two registers closely related to informal and spoken language, i.e. private letters and play-texts, promoted P-stranding, with sermons in the third position; all these registers are interactional. Over the course of the period studied, the frequency of P-stranding declined in most of the registers studied, pointing to the possible effectiveness of proscription condemning stranded prepositions (Yáñez-Bouza 2015a: 132–133). P-stranding has also been shown to be more prone to occur in spoken than in written registers in Present-day English (Biber, Johansson, Leech, Conrad, & Finegan 1999: Section 2.7.5.4; Hoffmann 2011: 76), and this turned out to be the case with the diachronic data. The quantitative results obtained could be seen to reflect the stylistic drifts attested by Biber and associates for the period (Yáñez-Bouza 2015a: 134–136). To give examples of further studies in this category, Smitterberg (2005) investigated the integration of the progressive form in nineteenth-century English. This study was based on the multi-genre CONCE corpus, and showed, among other things, that the construction was more common in women's than in men's private letters. Another corpus-based study of a single linguistic variable was Walker (2007) on the use of *you* vs. *thou* in Early Modern trial proceedings, witness depositions, and drama comedy included in the CED. Like Smitterberg (2005), this study highlights the effects of extralinguistic and linguistic factors for the developments. While previous studies of *you* and *thou* were mainly concerned with drama and were qualitative in nature, Walker (2007) extends the scope of the register traditionally considered in the study of the variable to include new speech-related

genres such as depositions and also uses frequency studies of occurrence patterns to support her conclusions.

3. What are the major methodological approaches that are used to analyze or account for register in historical linguistics?

Within the variationist studies framework (Weinreich, Labov, & Herzog 1968), register appears as a powerful factor to be considered when investigating the trends and rates of change. A pioneering work making use of this approach was Romaine (1982), in which the author correlated linguistic variation with extralinguistic factors, register being one of the main parameters observed. The work paved the way to the field of sociohistorical linguistics, one of the influential fields in historical linguistics today (note that this approach is in contrast to the way synchronic variationist studies have developed, with their focus typically being on controlled 'styles' of the 'sociolinguistic interview', as opposed to registers or naturally-occurring text; see Biber & Conrad 2009:264–265 for discussion). The popularity of the variationist framework was increased by the arrival of stratified electronic corpora compiled to comprise carefully selected texts representative of successive time periods and different (often contrasting) registers. Further advances in corpus linguistic techniques allowing automatic or semi-automatic normalization of spelling in early corpus texts, and linguistic and grammatical annotation of language data have further strengthened the position of variationist approaches to language change.

In English historical linguistics, recent research has focused on various processes of change, among them standardization, colloquialization, Americanization, grammaticalization and subjectification phenomena, processes of sociolinguistic and socio-pragmatic change, contact-related changes, and processes of change brought about by the spread of English around the world. The study of colloquialization phenomena, in particular, has proved a fruitful area (see, e.g., Mair 1997; Hundt & Mair 1999; Leech, Hundt, Mair, & Smith 2009). Over the past two decades, the study of early 'spoken interaction' has gained momentum, aiming at probing into the question of what past spoken language was like as no authentic audio-data exists before the late 1870s (Mair 2016). These processes are often studied by investigating changes in distributions of linguistic variants or constructions. Registers and genres have been considered to be important across the history of English, but they start to diversify and become better represented than before from the 1500s onward. This makes Early and Late Modern English periods particularly rewarding objects from the point of view of register and genre studies. In terms of register variation, the twentieth century, especially the first half of it, is

still among the periods in need of further representation in corpus material both regarding spoken and written data. What we already have are the 'extended Brown family' corpora, comprising one-million-word collections representative of varieties of English matching the original *Brown Corpus* (written American English from the 1960s) and *Lancaster-Oslo/Bergen Corpus* (LOB) (written British English from the 1960s). Among the extended Brown family corpora are *BLOB-1931 Corpus* (sampled from 1928–1934) and its counterpart *B-Brown-1931 Corpus* (both still in in-house use) (Hundt & Leech 2012; Leech & Smith 2005). There is currently a lively interest in the study of recent change (as of the 1960s onwards) and the role played by registers in it and in the period immediately preceding it. Similarly, the recent *British National Corpus 2014* (BNC2014) project can be expected to inspire further micro-diachronic research on register, genre and text type issues. The project has already resulted in the release of the *Spoken BNC2014* corpus (transcripts of recorded conversations from 2012–2016) and in the compilation of its counterpart, which will comprise written texts to be released in the near future.

The use of corpus evidence combined with the historical sociolinguistic and historical pragmatic approaches have inspired researchers to look into issues in grammatical change (morphology, syntax), and semantic and pragmatic phenomena. Most of these areas lend themselves to corpus linguistic search techniques (for a seminal paper, see Ball 1994). These studies showcase the interest that historical linguists have in pinning down the connections between the two extralinguistic variables, register and time (see, e.g., Hundt & Mair 1999). By surveying change proceeding within an apparent-time framework across successive synchronic subperiods and across various registers at different paces, some registers appear as more prone to accommodate change than others. Regarding quantification, Biber, Egbert, Gray, Oppliger, and Szmrecsanyi (2016; Biber 2012:12–17) distinguish two main approaches to the study of grammatical change, viz. variationist and text-linguistic. Within the variationist approach, it is the individual linguistic choices and what influences them that are in focus. In these investigations the unit of analysis is the occurrence or non-occurrence of each variant form, and the results report on the proportional preferences attested in the texts or registers. Within the text-linguistic approach, one focusses on the rates of occurrence in texts: the unit is each individual text, and the results report on how often a particular linguistic feature occurs in a text or register (Biber et al. 2016:356–357). In their study of the use of nominal modifiers of head nouns across a sub-corpus of the registers of personal letters, newspaper reportage, and science articles drawn from the ARCHER corpus, Biber et al. (2016) studied the patterns of variation in the use of constructions such as *the Communist Party chief*, *the Communist Party's chief* and *the chief of the Communist Party*, and demonstrated how the two methods yielded different but complementary descriptions of grammatical change in the data. It also became

clear that a single-register approach would have revealed misleading trends and that for the full picture of the development, both variationist and text-linguistic research designs were needed (Biber et al. 2016: 373–374).

According to Diller (2001), the histories of registers and genres are what compose the history of English. Each of these histories has its specific trajectory of change. While the histories of linguistic items and constructions have been a target of intensive research, the study of register and genre change is still a much less explored area. What has hampered and will keep on hampering research on historical registers is the elusive nature of evidence. Texts that were produced by less known or even unknown language users for everyday use were not considered of importance or worth being preserved (Kytö & Pahta 2012: 127–128; see, also, Smitterberg 2016: 192). This has had consequences to many of the registers of value for historical register analysis. On the other hand, genres and registers are valuable tools *per se* in helping one to group texts that have survived in haphazard or skewed ways into larger units for quantitative and qualitative analyses. They also highlight the possibility of internal register change (e.g. the emergence of new sub-registers) across the centuries or point to interruptions in the continuity of a register in the course of time (Kytö & Pahta 2012: 125).

4. What does a typical register study look like in historical linguistics?

As indicated above, it is common in studies in historical linguistics to approach registers as external factors playing a role in change. It is convenient and usually also justified to take the register classification adopted for an electronic corpus as a starting point and look into the frequency distributions of linguistic variant forms across the registers and time. Rises or dips in frequencies of variant forms are interpreted as incipient change or change in progress or completed, and the associated registers are considered as factors or environments contributing to change by either promoting or retarding it. Depending on the occurrence of the variant forms across registers, especially if occurrences remain scanty, it may turn out necessary to pool together related registers for larger units. In many, if not most, studies based on historical corpora, the coding adopted for the corpus files has been used unchanged when reporting on distributions of variant forms. More rarely, register definitions have been explored further, as in, for instance, Yáñez-Bouza (2015b), where the similarities and differences between diaries and journals are investigated in a systematic way and where it is argued that there are grounds for re-classifying the texts included in one and the same diaries/journals register in ARCHER in two separate registers.

A recent large-scale empirical investigation of linguistic change making full-fledged use of the register parameter is Rodríguez-Puente (2019) on the variation and development of phrasal verbs in the history of English (I am grateful to the author for permission to use the proofs of the book as basis for this account of the study). This monograph is of major interest to historical register studies for a number of reasons. First, registers as such have been shown to be of major importance in the development of phrasal verbs across the history of English; secondly, the author has made innovative use of a great number of electronic resources covering a wide spectrum of texts from the Early Modern period onward; and thirdly, as most previous research has shown that the frequency of phrasal verbs in English has increased considerably from Early Modern English (EModE) times (see, e.g., Spasov 1966:125; Pelli 1976:102; Martin 1990; Wild 2010:227; Diemer 2014, as referred to in Rodríguez-Puente 2019:1), it is via the lens of various registers that the present study aims at assessing the varying register-specific rates of change in the development. The ten genres included in the systematic study are diaries and journals, personal letters, drama and fiction, news, medicine and science, sermons, and trial proceedings. Both quantitative and qualitative analyses are presented, with the aim of identifying the paths of change relating to the registers on the formal vs. informal and speech vs. writing continua along the lines of the multi-dimensional models introduced by Biber in his (1988) study and subsequent work (e.g., Biber 2001; Biber & Finegan 1989, 1992, 1997, 2001; see Rodríguez-Puente 2019:12). In addition to the register factor, the other external factors taken into account include subject matter, structural properties of a text (e.g., repetitive, sketchy notes in diaries and reports of medical or scientific experiments), related changing stylistic conventions, and individual author preferences.

The period covered by the study extends from 1650 to 1990, and the main bulk of evidence is drawn from ARCHER and a sample of the *Old Bailey Corpus* (OBC, version 1.0, see Huber 2007), with additional information drawn primarily from the ICE-GB (a section of the *International Corpus of English Great Britain*) and also from the *Helsinki Corpus*, the *Corpus of Late Modern English Texts Extended Version* (CLMETEV), the *Cambridge International Corpus*, the *British National Corpus* (BNC), the database of examples of the *Oxford English Dictionary*, the World Wide Web, and a number of other historical and present-day corpora of American and British English (Rodríguez-Puente 2019: Chapter 2). The focus is on British English, but comparisons are made with other regional varieties, especially with American English.

All in all, the primary dataset comprises over 13,200 instances of phrasal verbs collected from ARCHER and the OBC, with illustrative examples taken from other sources (Rodríguez-Puente 2019:9). The aim of the study was to

investigate the morphological, semantic and syntactic characteristics of phrasal verbs across the registers represented in the material and to test whether the use of phrasal verbs could be related to spoken, colloquial registers, in particular. Interestingly, the author seeks to pin down the role played by the processes of grammaticalisation, lexicalisation and idiomatisation. The monograph presenting the study comprises six chapters.

In her introduction to the study, Rodríguez-Puente explains that phrasal verbs are characteristic of Germanic languages and comprise expressions such as *give up*, *find out*, *fall out*, and *put down*. For some verbs their meaning can be deduced from the union of the two (or more) elements (e.g. *give up* and *find out*) but other combinations can result in non-compositional idiom-like expressions that are difficult to make sense of as such (e.g. *fall out* 'argue' and *put down* 'criticize') (Rodríguez-Puente 2019: Section 1.1). In Rodríguez-Puente's view, phrasal verbs are "combinations of a verb and a post-verbal particle, which function as relatively unitary structures lexically and semantically as, for example, *put up*" (Rodríguez-Puente 2019: 2–3). What is an essential criterion here is the nature of the post-verbal particle, which with phrasal verbs is an adverb (and not a preposition); with prepositional verbs (e.g., *look after*) the particle is a preposition and with phrasal-prepositional verbs (e.g., *break in on*) the particle is a combination of an adverb and a preposition (for terminology, the author refers to Aarts 1992: 89; Claridge 2000: 46; Cappelle 2007: 41–43) (Rodríguez-Puente 2019: 2). The author deals with the status of phrasal verbs in Chapter 3 and prefers to view phrasal verbs as a gradable category, aligning with those previous studies where meanings of phrasal verbs are best understood within a scale ranging from literal to idiomatic. At the same time, she also argues for a more fine-grained characterisation which also includes reiterative (e.g., *rise up*), emphatic (e.g., *wrap up*), and metaphorical combinations (*throw away a fortune*).

As is advisable in corpus linguistic studies, the corpora and other electronic resources used for the study are introduced in some detail, with attention paid to their suitability for the research envisaged. In her study on phrasal verbs, Rodríguez-Puente deals with the advantages and limitations of ARCHER for the aims of her study. As ARCHER was created to enable the analysis of historical change in the range of written and speech-based registers of English from 1650 to the present (Biber & Finegan 1997: 255), it is well suited for the purposes of a study of the history of phrasal verbs. ARCHER version 3.1 (2006) used for the study comprised close to 1.8 million words distributed over 955 files across a 50-year sub-periodization plan, and the British component in this version is divided into seven fifty-year sub-periods, with eight different genres represented (drama, fiction, journals/diaries, letters, medicine, news, science, and sermons), totaling more than 1.2 million words distributed over 674 files (Rodríguez-Puente 2019:

Section 2.2.1). Access to texts representative of informal and colloquial styles is a significant advantage, considering that in Present-day English phrasal verbs have been associated with these styles. Moreover, the previous multivariate analyses carried out on the material have already shown that registers may vary over time and that conventions associated with certain present-day registers may not have held for the early periods (see, e.g., Biber 1988, 2001; Biber & Finegan 1989, 1992, 1997, 2001) (for discussion, see Rodríguez-Puente 2019: Section 2.2.1.1). Among the limitations of ARCHER the author mentions lack of full representation of American English, inconsistencies and duplication of texts in the files included in this particular version of the corpus (amended in the subsequent versions), and lack of speech-based texts proper (while other speech-related registers showing varying degrees of speech-likeness such as diaries, drama, letters, and sermons are represented) (Rodríguez-Puente 2019: Section 2.2.1.2).

To supplement the representation of past speech, a sample of over 450,000 words of trial records was drawn from the 14-million-word OBC from 1720 to 1913. These printed records were based on the shorthand notes taken down by scribes in the courtroom, making the texts as reliable a source as can be expected under the circumstances, even considering the variation caused by changes in scribal and printing practices across the period. The selection of texts sampled from the OBC for the present study of phrasal verbs only contains direct speech. Also, regarding a study on the development of phrasal verbs, recording features such as omission of slips of tongue, false starts, discourse particles, or contracted forms would not affect the results in any substantial way (Rodríguez-Puente 2019: Section 2.3.7).

To provide evidence of the use of phrasal verbs in the twentieth century beyond the OBC, Rodríguez-Puente turned to the legal cross-examinations of the British component of the ICE-GB, extending from 1990 to 1993. This meant leaving a gap in the textual representation and corpus data (the 1920s to 1970s) but to make comparisons more reliable, it was deemed justified to focus on the sub-genre of legal cross-examinations found among the public dialogues of the spoken section of the ICE-GB, 10 texts totaling some 21,352 words. To supplement the picture for Present-day English, a variety of other sources were consulted, among them the BNC, World Wide Web, and the database of the *Oxford English Dictionary* covering 1000 years of the history of English (Rodríguez-Puente 2019: 24–26).

Before proceeding to account for her search methodology, Rodríguez-Puente gives careful consideration to the different registers included in the study, paving the way for the reclassification of the ARCHER texts (see Section 2.3). The main concern here is the changing conventions of registers and genres over time and the special characteristics of the ARCHER and OBC texts. Only some examples of the points raised can be given here. Along the lines indicated in Yáñez-Bouza (2015b), with reference made to Biber and Conrad (2009) for analyzing register

variation in the light of the relation between participants, channel/mode, production circumstances, setting, communicative purpose(s), and topic, Rodríguez-Puente treats diaries and journals as two albeit related yet separate genres. Further, letter writing is subject to register variation owing to writers' varying social background and their intimate or distant relationship with the addressee. Letters have indeed been shown to vary as a genre across time, especially from the eighteenth century to modern times (for multivariate analyses, see Biber 2001; Biber & Finegan 1989, 1997). Other registers that have gone through major changes when compared with their present-day counterparts include news reportage, scientific and medical discourse, sermons, and trial proceedings.

For the purposes of her study and drawing on terminology from Culpeper and Kytö (2010:17), Rodríguez-Puente reclassified the ARCHER and OBC texts into two major groups: writing-based and writing-purposed texts (fiction, journals, news, medicine, and science) and speech-related texts (diaries, drama, letters, sermons, and trial proceedings). Three sub-groups were distinguished in the latter category: speech-like texts (letters and diaries), speech-based texts (trial proceedings) and speech-purposed (drama and sermons). Speech-like texts are oral or colloquial in the sense that they reflect features of communicative immediacy, speech-based texts are based on actual authentic speech events relayed via shorthand or other types of note-taking practices, and speech-purposed texts have been designed for oral articulation (Rodríguez-Puente 2019: Section 2.3.8; cf. Koch & Oesterreicher 1985). The division of texts into writing-based/writing-purposed and speech-related texts implies interrelated and overlapping rather than clear-cut categories, and this also holds for the level of formality of texts: towards the informal end Rodríguez-Puente placed diaries, drama, letters, and dialogues in trial proceedings, and towards the formal end sermons, medicine, and science, which left fiction, journals, and news in the middle ground.

To extract the instances of phrasal verbs from the corpora, concordances were created for the particles included in the study, and the relevant examples of phrasal verb uses were identified manually. The 34 particles searched for were drawn from Claridge (2000:46), which was based on items included in previous studies on the topic. Once the output files had been screened, the relevant instances were stored in Microsoft Excel databases to be coded for linguistic and extralinguistic factors. For comparisons, the differences in text length across the texts and registers were remedied by normalizing the raw figures per 10,000 words. Statistical tests were used to verify the statistical significance of the variables where appropriate.

The range and versatility of corpora and the 34 particles included in searches produced a rich array of results, reported in Chapter 6 of the book. Overall, while there was considerable variation across both the registers and the sub-periods distinguished, phrasal verbs proved significantly more frequent in trial proceedings

than in the other registers, and they were also more common in journals, diaries, and fiction than in science, medicine, and news. A mixture of tendencies seemed to characterize register-specific tendencies. There were no patterns for the very high and very low frequencies scored by speech-related text types and by writing-based and writing-purposed genres. The degree of (in)formality revealed similar variable tendencies.

When zooming to the use and development of phrasal verbs in individual genres, Rodríguez-Puente turns to factors beyond the relationship of a text to spoken language and level of (in)formality. Among these were the subject matter, the role played by linguistic correctness, and authorial preferences, e.g., the use of Latinate verbs instead of phrasal verbs. The type-token ratio (TTR) values and the results obtained in previous corpus studies are also brought into discussion for comparisons. In what follows, the main results of the present study and the author's interpretations of them are surveyed (for details, see Rodríguez-Puente 2019: Chapters 5–6).

Diaries and journals offered further evidence on the usefulness of treating them as separate genres (cf. Yáñez-Bouza 2015a, 2015b): while diaries displayed an increasing use of phrasal verbs, characteristic of the so-called popular written genres, from the eighteenth century to the present day, the use of phrasal verbs in journals was similar to the use displayed by medical and scientific writing and other formal written registers. In quantitative terms the results were not statistically significant, however, but Rodríguez-Puente suggests a number of explanations to account for the obviously random variation in the data across the period studied. Diaries tend to be written in freer style than journals, reflecting the ways in which the subject matter and the preferences of writers formed by their sociocultural and -educational background influence the contents (Rodríguez-Puente 2019: Section 6.2.1).

In personal letters, phrasal verbs were used to a much lesser extent than in other informal registers up until the 1800s, with increasing rates from the nineteenth century onwards (Rodríguez-Puente 2019: Section 6.2.2). Rodríguez-Puente ascribes this to the subject matter of the letters (few descriptions of places, trips, and travels inducing the use of phrasal verbs) and shift towards more oral styles, early letters having been more literate in character (with reference made to Hiltunen 1994: 137, who had already observed this in his analysis of the letters in the *Helsinki Corpus*). The increase of phrasal verbs took place relatively continuously, with two reversal periods bucking the trend. There were also writers who seemed to have consciously avoided the use of phrasal verbs for one-word Romance equivalents.

Fiction and drama show rising frequencies across the period, with higher figures for fiction in general apart from some sub-periods (Rodríguez-Puente 2019:

Section 6.2.3). From the 1750s onwards, fiction takes off with steadily increasing rates, leaving drama behind, with the early 1800s as the take-off point. Rodríguez-Puente suggests that the low rates of occurrence for phrasal verbs in these genres during the eighteenth century might be accounted for by changes in the presentation conventions of the time and refers to the results obtained in Biber and Finegan (1989, 1997) which pointed to a dramatic change in the linguistic characteristics of fiction and other genres with a drift towards more oral styles over the past 400 years. Further support is found in McIntosh (1998: 1, 23), who relates the change towards more 'polite' and more 'written' (less 'oral') style in eighteenth-century fiction to prescriptivism which encouraged formality and precision, discouraging oral features, such as, as Rodríguez-Puente points out, phrasal verbs. The frequencies of such features increase as of the 1800s onwards with change towards more involved styles (Biber & Finegan 1989: 507).

News reportage and newspapers overall differ from other popular written genres regarding their particular stylistic and linguistic conventions, which largely go back to commercial and marketing aims. Newspapers have been subject to major change across their history along with advances in communication technologies, and this can be seen in changes in the use of phrasal verbs as well. Relatively frequent use was made of them in the earliest sub-periods, but frequencies fall up until the 1800s; after a rise (1800–1850) and a fall (1850–1900), they increased again after the 1900s. Several previous studies have recorded the increased colloquialization trend towards the end of the twentieth century (with reference made to Biber & Finegan 1997; Mair 1997; Rühlemann & Hilpert 2017). Again, other factors than the level of orality seem to be at play (Rodríguez-Puente 2019: Section 6.2.4).

In medicine and science, apart from a rise in medicine in the 1700–1749 sub-period and a sudden increase in science in the 1800–1849 period, the use of phrasal verbs has declined over time, especially from the 1850s onwards. While the overall decline tallies well with expository registers becoming more literate as shown by, e.g., Biber and Finegan (1997) and Atkinson (1999), other factors will need to be investigated to account for fluctuation in the rates of change. One factor causing rises in the figures was the diary form in which some of the individual scientific texts were rendered in the 1800–1849 sub-period. These texts could be sketchy and repetitive in nature, with repeated use made of certain phrasal verbs. Another factor causing fluctuation in figures obtained for individual texts was subject matter, e.g., the description of various movements carried out during the examination and treatment of a patient. Of these two related genres, medical writing has been more variable over time, in fact, displaying the highest rates of outliers in ARCHER (Rodríguez-Puente 2019: Section 6.2.5).

Sermons are characterized by an increasing use of phrasal verbs after preachers had turned to a plainer style from the 1700s onwards, a result in line with previous studies (e.g., Hiltunen 1994; Claridge 2000). In fact, phrasal verbs are more frequent in sermons in ARCHER than in letters between 1650 and 1850. Evidence of idiolectal preferences and subject matter can also be found in individual sermons. Also, a likely source for the use of certain combinations such as *bring forth*, *lay down*, and *lift up* was biblical language (Rodríguez-Puente 2019: Section 6.2.6).

Trial proceedings display the highest figures for phrasal verbs in all the genres in all the sub-periods, apart from 1850–1913 in the proceedings (in this period journals scored higher). As these texts are based on authentic speech, Rodríguez-Puente concludes that "phrasal verbs were typical of the spoken language as early as the eighteenth and nineteenth centuries" (Rodríguez-Puente 2019: 264–265). The finding is supported by previous research (e.g., Claridge 2000: 190–192; Thim 2012: 233–46). Yet, the frequencies of phrasal verbs show an overall decline as of the 1750s onwards, after a minor increase in the first half of the eighteenth century. The decline is confirmed by the figures obtained for the ICE-GB data in legal cross-examinations from the 1990s, although Rodríguez-Puente does draw attention to the differences in the recording styles between the two data sources (the OBC material containing lengthy testimonies relaying earlier conversations and the ICE-GB material consisting of questions and 'yes' or 'no' answers). She also postulates a process of 'colloquialization' that might account for the decline in the use of a colloquial feature such as phrasal verbs, making reference to Culpeper and Kytö (2010: 103–141), on linguistic developments in Early Modern English trial proceedings bucking "the trend towards becoming more speech-like" (2010: 137) as the business of the courtroom was shifting into the direction of formal and formalized registers (2010: 139). Further support is lent by Widlitzki and Huber (2016) indicating a reduction of bad language and insults in the OBC across time.

On the whole, Rodríguez-Puente finds that her study is the first to assess the effect of phrasal verbs on various genres of ARCHER, since Biber (2003), where this linguistic feature was found to load positively on Dimension 1. In her view the fact that her register-specific findings on the development of phrasal verbs follow the general trends attested in previous multivariate analyses is evidence against the claims that phrasal verbs were stylistically neutral in Early Modern English rather than associated with spoken language (Thim 2006, 2012) although further research is needed on, e.g., the *Corpus of English Dialogues 1560–1760*. As for the Late Modern English period, Rodríguez-Puente finds that she has presented conclusive evidence of phrasal verbs being a feature of spoken language of the time, with increase in their use coinciding with colloquialization of popular written texts, and, conversely, with decrease in their use coinciding with the decolloquialization of specialized written genres, such as medicine, science, and

legal cross-examinations. As to the question of why phrasal verbs became associated with spoken language registers, a number of explanatory circumstances are suggested, among them the avoidance of Latinate and Romance one-word counterparts. Another factor comprised increased stress possibilities, especially along with the emergence of emphatic particles used to reinforce the verb meaning (e.g., *boil up*, *splash up*) in the Late Modern period and notably in speech-related registers; these particles were often criticized as redundant in grammars in the 1700s and 1800s (Wild 2010). Further conditioning factors included subject matter, form of writing, and idiolectal preferences.

5. What are the most promising areas of future register research in historical linguistics?

Registers have been shown to act as powerful vehicles promoting or retarding language change, or even contributing to stability. Overall, considering recent paradigm shifts in language studies, context – whether linguistic, social or cultural – has been gaining in importance, making register studies more central to the study of change. Registers are an important component in what has been referred to as the 'pragmatic turn' by which is meant that "[v]ariation through space, social group and time, multi-disciplinarity, and multiculturalism have all come to play center stage" (Traugott 2008: 207–208; for discussion, see Taavitsainen et al. 2014). Research aimed at disentangling the ways in which various registers have developed across time and influenced change is thus of major interest. The question of the relationships between registers and sub-registers in diachronic perspective is an under-researched area and in need of systematic work with attention paid to the interplay of extralinguistic and linguistic features. Internal change in registers across time is another area which would merit further rigorous attention (cf. the progress made in this area regarding, e.g., scientific and medical writing): general genre and register labels are often used irrespective of what the unit referred to actually comprises in reality. Further, the question of how to attend to the subject matter of texts as a factor interacting with register variation would need further consideration. Recent developments in semantic tagging, for instance the Samuels Tagger, should be of help in this respect, combined with, e.g., multi-dimensional approaches. In terms of the fundamentals of change and especially the important role played by everyday spoken language, developing ways in which one could have access to registers conveying glimpses at past spoken interaction would be very valuable.

Regarding methodology, register studies would benefit from approaches where quantitative and qualitative approaches go hand in hand and where the

synergy effects of multi-dimensional, variationist and text-linguistic analyses have been attended to. To enable progress in register studies, labor-intensive work is required regarding both the extralinguistic and linguistic annotation of the currently available corpora and the compilation of new resources containing fresh material. Many of the corpora currently available to researchers only have limited information on the extralinguistic parameters of texts or registers and lack grammatical and other linguistic annotation. New corpora with enhanced register representation and linguistic annotation would add significantly to the potential of register studies. This is a very important aspect especially considering the increasing interest in the innovative use made of 'big data' among, e.g., historians, social historians, and those interested in gender studies, mass media and many other areas. Some of the areas in English linguistics that have recently emerged as foci of attention and that are of interest to further work on register studies are described below.

Indeed, big data and corpus linguistic techniques meet fruitfully in what has recently been termed as 'digital humanities'. This is an area full of promise for interdisciplinary research on historical registers bringing together linguistics, literary history, social history, and cultural studies. For instance, in their recent monograph, McEnery and Baker (2017) explore representation of prostitution in the 16th century England by using data drawn from *Early English Books Online* (EEBO) and by combining the study of socio-cultural history with corpus and computational linguistics. When investigating usage patterns of words such as *whore*, the authors consider the influence of genres and discover that it is in plays, in particular, that the word is used as an insult (McEnery & Baker 2017: 181). Work has certainly been done on EEBO and other register and subregister-rich electronic resources such as *Literature Online* (LION) and *Google Books*, but these collections still hold potential which has not been exhausted yet.

Considering recent topical approaches, the framework of "enregisterment" applied in, for instance, sociocultural linguistics and computer-mediated communication provides ample room for historical register studies. For instance, following Agha (2003, 2006), Hoffman & Kytö (2017) recognize that by way of enregisterment, language users employ specific linguistic features to index certain cultural values; thus a subgenre such as early American-Swedish cookbooks was found to display recipe names in 'heritage Swedish' (*Beef à la Lindstrom*) or standard Swedish (*Biff à la Lindström*) that were linked to the Swedish-American community (Hoffman & Kytö 2017: 268). The various genres representing writing in public and private spheres in heritage communities with various types of multilingual settings offer sociolinguistic environments particularly suited for observing the evolving dynamics of enregisterment (cf. Jaffe 2015: 562, on enregisterment

being best surveyed "as a process that can only be known retroactively, defined by its outcome").

To return to an approach already mentioned above, there is significant unexplored potential for historical register studies in the 'communities of practice' framework (Wenger 1998). According to Meyerhoff (2002), communities of practice are characterized by mutual engagement and jointly-negotiated enterprises with activities that have resulted in shared linguistic repertoires (cf. Hoffman & Kytö 2018: 46). Historical registers have developed as a response to various communicative needs by interacting members of communities of practice, and they have played a crucial role across the centuries for e.g. networks of letter writers, communities of scribes and printers, and groups of professionals such as administrators and scientists (for discussion, see Jucker & Kopaczyk 2013). This is again an area full of promise also for languages other than English.

Historical corpora have so far been traditionally based on early imprints or printed editions that reflect a range of varying editorial practices often making it difficult to assess the reliability of corpus texts for various types of linguistic investigations. An area in need of urgent further work is manuscript studies, especially regarding the Early Modern and Late Modern English periods when register representation is expanding and diversifying, allowing one to make comparisons with printed sources. Only very little has been done so far in this area to create reliable data sources for linguistic community. Electronic text editions, faithful to original manuscript text and structured as language corpora would provide access to new data sources whether based on documentary or literary works. Examples of such corpus-linguistically oriented digital documentary editions include Meurman-Solin (2007) on early Scottish correspondence, Kytö, Grund, and Walker (2011) on witness depositions drawn from ecclesiastical and criminal courts located in different areas of early Modern England, and Marttila (2014) on six fifteenth-century culinary recipe collections, the Potage Dyvers family. It would mean a major advance to historical register studies, if further manuscript sources representative of various registers were made available in digitized text editions. For instance, there is an increasing interest in the so-called ego documents that tend to allow room for spontaneous and unsupervised language use for authors. In addition to the extant letter corpora (mostly based on traditional text editions), one would welcome personal letters, diaries and journals based on manuscript sources.

Another fascinating area of promise for register studies has been the growing body of work on the history of transplanted varieties of English. New circumstances in the new country have meant novel needs for migrants regarding the kind of texts of interest to them. Examples of resources displaying register representation from early colonies include the *Corpus of Early Ontario English* (CONTE) with its pre-Confereration section (CONTE-pC) (Dollinger 2006,

2008: 99–120). This resource spans the period from the earliest Ontarian English texts to the end of the nineteenth century and comprises some 225,000 words; the pre-Confederation section (pC) included in the corpus spans the period 1776–1850 and comprises some 125,000 words in three genres, i.e., newspaper texts, diaries, and letters. The compilation principles and the design of the corpus make the material comparable to other language corpora, e.g., ARCHER. The relatively small size of the corpus leaves room for further compilation work.

Another corpus devoted to an early variety of World Englishes, the *Corpus of Oz Early English* (COOEE), covers the period 1788–1900 and comprises texts representative of language use in Australia, New Zealand, and the Norfolk Island, totaling two million words (Fritz 2007). The definition for registers and genres was influenced by Biber and associates, the main effort invested in covering the range rather than proportions of variation in the material (Biber, Conrad, & Reppen 1998; for discussion, see Fritz 2007: 70). Four registers are covered in the corpus (capital initials are used in the register and genre labels adopted): the Public Written Register, the Private Written Register, the Speech-based Register, and the Register of Government English. Of these the Public Written Register comprises Memoirs, Newspapers & Broadsides, Narratives, Official Correspondence, Reports, and Verse while the Private Written Register comprises diaries and personal correspondence. These two registers cover 75% of the material, leaving Speech-based Register and Government English in the minority mainly owing to shortage of suitable corpus texts. The former comprises Minutes, Plays and Speeches, and the latter Imperial Correspondence, Legal English, and Petitions and Proclamations. Such rigorous register and text type classification is very attractive in that it, among other things, enables comparisons with contemporaneous material representative genre and register variation in the mother country. Also, while the corpus proper comprises two million words, the compiler has eight million words at hand as possible reference material (Fritz 2007: 70–71).

As for American English, the 400-million-word *Corpus of Historical American English* (COHA) is representative of four registers, i.e., fiction, magazine, newspaper, and non-fiction/academic books (the term 'register' is used here in the wide sense given in Section 1). However, as COHA only contains texts from the 1810s onwards, it leaves the seventeenth and eighteenth centuries without representation. Yet, the early New England colonies offer a wide range of texts representative of utilitarian writing including ego documents such as letters, diaries and travel journals, and also history writing, sermons, witness depositions and other court records (for discussion, see Kytö 1991: 30–37). Also, while no drama texts were produced in the early colonies (the first theatre was opened only in Virginia in the early 1700s), there were certain colony-specific genres such as descriptions of the flora and fauna new to the settlers, and 'Memorable Providences', where

incidents purportedly revelatory of God's will and aspirations were reported on. The Southern colonies present their own register profile, which remains necessarily less versatile owing to the lack of general interest in education. Comparisons with register variation in the early colonies and the mother country would be of major interest regarding the formative years of American English.

From a bird's-eye view, registers, past and present, are about formality and informality, both terms which have inspired plenty of interesting work in cultural history. In particular, the notion of informalization has been found applicable to developments in European social and cultural history since the later Middle Ages (e.g., Elias 1976/1939) and can reasonably be expected to promote further register research on linguistic features such as the grammaticalization of pronominal address systems and other politeness phenomena in various European languages. Overall, linguistic developments associated with informalization and democratization trends (cf. Fairclough 1992; Fairclough & Wodak 1997) are of prime interest to historical register studies. This also holds for socio-semiotic approaches, which highlight the need to study forms of writing as linguistic or semiotic practices. These forms of writing have evolved functionally to serve the theoretical and practical needs in social institutions (see, e.g., Halliday & Martin 1993). Such perspectives offer exciting prospects for learning more about register variation in the history of not only English, but also other languages.

Corpora and other electronic resources

ARCHER, see *A Representative Corpus of Historical English Registers*
B-Brown-1931 Corpus
BLOB-1931 Corpus
British National Corpus (BNC)
British National Corpus 2014 (BNC2014)
Brown Corpus (The Standard Corpus of Present-Day Edited American English)
Cambridge International Corpus
Corpus of Early English Correspondence (CEEC)
Corpus of Early English Medical Writing (CEEM)
Corpus of Early Ontario English (CONTE)
A Corpus of English Dialogues 1560–1760 (CED)
Corpus of Historical American English (COHA)
Corpus of Late Modern English Texts Extended Version (CLMETEV)
A Corpus of Nineteenth-century English (CONCE)
Corpus of Oz Early English (COOEE)
Early English Books Online (EEBO)
Google Books
Helsinki Corpus (Helsinki Corpus of English Texts)

International Corpus of English Great Britain (ICE-GB)
Lancaster-Oslo/Bergen Corpus (LOB)
Literature Online (LION)
Old Bailey Corpus (OBC)
A Representative Corpus of Historical English Registers (ARCHER)
Spoken BNC2014
Zurich Newspaper English (ZEN)

References

Aarts, B. (1992). *Small clauses in English: The non-verbal types*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110861457

Agha, A. (2003). The social life of cultural value. *Language & Communication*, 23, 231–273. https://doi.org/10.1016/S0271-5309(03)00012-0

Agha, A. (2006). *Language and social relations* (Studies in the Social Life and Cultural Foundations of Language 24). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511618284

Arroyo, B., Luis, J., & Schulte, K. (2017). Competing modal periphrases in Spanish between the 16th and 18th centuries: A diachronic variationist approach. *Diachronica*, 34, 1–39. https://doi.org/10.1075/dia.34.1.01bla

Atkinson, D. (1992). The evolution of medical research writing from 1735 to 1985: The case of the Edinburgh Medical Journal. *Applied Linguistics*, 13, 337–374. https://doi.org/10.1093/applin/13.4.337

Atkinson, D. (1996). The *Philosophical Transactions* of the Royal Society of London, 1675–1975: A sociohistorical discourse analysis. *Language in Society*, 2(3), 333–371. https://doi.org/10.1017/S0047404500019205

Atkinson, D. (1999). *Scientific discourse in sociohistorical context*. The Philosophical Transactions of the Royal Society of London, 1675–1975. Mahwah, NJ: Lawrence Erlbaum Associates.

Bakhtin, M. M. (1986 /1953). *Speech genres and other late essays*. Austin, TX: University of Texas Press.

Ball, C. N. (1994). Automated text analysis: Cautionary tales. *Literary and Linguistic Computing*, 9, 295–302. https://doi.org/10.1093/llc/9.4.295

Bazerman, C., & Paradis, J. (1991). *Textual dynamics of the professions: Historical and contemporary studies of writing in professional communities*. Madison, WI: University of Wisconsin Press.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

Biber, D. (2001). Dimensions of variation among eighteenth-century speech-based and written registers. In H.-J. Diller & M. Görlach (Eds.), *Towards a history of English as a history of genres* (pp. 89–109). Heidelberg: Universitätsverlag C. Winter.

Biber, D. (2003). Variation among university spoken and written genres: A new multi-dimensional analysis. In P. Lestyna & C. F. Meyer (Eds.), *Corpus analysis. Language structure and language use* (pp. 47–70). Amsterdam: Rodopi.

Biber, D. (2012). Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory*, 8(1), 9–37. https://doi.org/10.1515/cllt-2012-0002

Biber, D., & Burges, J. (2000). Historical change in the language use of women and men: Gender differences in dramatic dialogue. *Journal of English Linguistics*, 28(1), 21–37. https://doi.org/10.1177/00754240022004857

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511814358

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics – Investigating language structure and use*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

Biber, D., Egbert, J., Gray, B., Oppliger, R., & Szmrecsanyi, B. (2016). Variationist versus text-linguistic approaches to grammatical change in English: Nominal modifiers of head nouns. In M. Kytö & P. Pahta (Eds.), *The Cambridge handbook of English historical linguistics* (pp. 351–375). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139600231.022

Biber, D., & Finegan, E. (1989). Drift and the evolution of English style: A history of three genres. *Language*, 65(3), 487–517. https://doi.org/10.2307/415220

Biber, D., & Finegan, E. (1992). The linguistic evolution of five written and speech-based English genres from the 17th to the 20th centuries. In M. Rissanen, O. Ihalainen, T. Nevalainen & I. Taavitsainen (Eds.), *History of Englishes: New methods and interpretations in historical linguistics* (pp. 688–704). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110877007.688

Biber, D., & Finegan, E. (1997). Diachronic relations among speech-based and written registers in English. In T. Nevalainen & L. Kahlas-Tarkka (Eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen* (pp. 253–275). Helsinki: Société Néophilologique.

Biber, D., & Finegan, E. (2001). Diachronic relations among speech-based and written registers in English. In S. Conrad & D. Biber (Eds.), *Variation in English: Multi-dimensional studies* (pp. 66–83). London: Pearson Education.

Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics*, 15(2), 223–250. https://doi.org/10.1017/S1360674311000025

Biber, D., & Gray, B. (2013). Being specific about historical change: The influence of sub-register. *Journal of English Linguistics*, 41(2), 104–134. https://doi.org/10.1177/0075424212472509

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.

Cappelle, B. (2007). When "wee wretched words" wield weight: The impact of verbal particles on transitivity. In M. Nenonen & S. Niemi (Eds.), *Collocations and idioms 1: Papers from the First Nordic Conference on Syntactic Freezes, Joensuu, Finland, 19–20 May 2006* (pp. 41–54). Joensuu: University of Joensuu.

Claridge, C. (2000). *Multi-word verbs in Early Modern English: A corpus-based study*. Amsterdam: Rodopi. https://doi.org/10.1163/9789004333840

Claridge, C. (2012). Linguistic levels: Styles, registers, genres, text types. In A. Bergs & L. J. Brinton (Eds.), *English historical linguistics: An international handbook* (Vol.1, 237–253). Berlin: Walter de Gruyter.

Culpeper, J., & Kytö, M. (2010). *Early Modern English dialogues: Spoken interaction as writing*. Cambridge: Cambridge University Press.

Diemer, S. (2014). Closing the gap – the development of verb-particle combinations in English between 1810 & 1960. In Y. Iyeiri & J. Smith (Eds.), *Studies in Middle and Modern English: Historical change* (pp. 41–57). Osaka: Osaka Books.

Diller, H.-J. (2001). Genre in linguistic and related discourses. In H.-J. Diller & M. Görlach (Eds.), *Towards a history of English as a history of genres* (pp. 3–43). Heidelberg: C. Winter.

Dollinger, S. (2006). Oh Canada! Towards the Corpus of Early Ontario English. In A. Renouf & A. Kehoe (Eds.), *The changing face of corpus linguistics* (Language and Computers: Studies in Practical Linguistics 55) (pp. 7–25). Amsterdam: Rodopi. https://doi.org/10.1163/9789401201797_004

Dollinger, S. (2008). *New-dialect formation in Canada: Evidence from the English modal auxiliaries*. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.97

Elias, N. (1976 /1939). *Über den Prozeß der Zivilisation: Soziogenetische und psychogenetische Untersuchungen* (2 vols, 4th ed.). Frankfurt: Suhrkamp. [*The civilizing process: Sociogenetic and psychogenetic investigations* (revised edition); translated by E. Jephcott with some notes and corrections by the author; edited by E. Dunning, J. Goudsblom & S. Mennell. Malden/Oxford: Blackwell, 2000].

Ellegård, H. A. (1953). *The auxiliary* do. *The establishment and regulation of its use in English* (Gothenburg Studies in English 2). Stockholm: Almqvist & Wiksell.

Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. In T. A. van Dijk (Ed.), *Discourse as social interaction* (pp. 258–284). London: Sage.

Fowler, A. (1982). *Kinds of literature: An introduction to the theory of genres and modes*. Oxford: Clarendon Press.

Fritz, C. (2007). *From early English in Australia to Australian English 1788–1900*. Frankfurt: Peter Lang.

Geisler, G. (2002). Investigating register variation in nineteenth-century English: A multi-dimensional comparison. In R. Reppen, S. M. Fitzmaurice & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 249–271). Amsterdam: John Benjamins. https://doi.org/10.1075/scl.9.17gei

Geisler, G. (2003). Gender-based variation in nineteenth-century English letter-writing. In P. Leistyna & C. F. Meyer (Eds.), *Corpus analysis: Language structure and language use* (pp. 87–106). Amsterdam: Rodopi.

Görlach, M. (2004). *Text types and the history of English*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197167

Halliday, M. A. K. (1988). On the language of physical science. In M. Ghadessy (Ed.), *Registers of written English: Situational factors and linguistic features* (pp. 162–178). London: Printer Publishers.

Halliday, M. A. K., & Martin, R. J. (Eds.) (1993). *Writing science: Literacy and discursive power*. London: Falmer Press.

Hauser, S., Roth, K. S., & Kleinberger, U. (Eds.) (2014). *Musterwandel – Sortenwandel: Aktuelle Tendenzen der diachronen Text(sorten)linguistik*. Bern: Peter Lang. https://doi.org/10.3726/978-3-0351-0668-8

Hiltunen, R. (1994). Phrasal verbs in Early Modern English: Notes on lexis and style. In D. Kastovsky (Ed.), *Studies in Early Modern English* (pp. 129–140). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110879599.129

Hoffman, A., & Kytö, M. (2017). The linguistic landscapes of Swedish heritage cookbooks in the American midwest, 1895–2005. *Studia Neophilologica*, 89(2), 261–286. https://doi.org/10.1080/00393274.2017.1301783

Hoffman, A., & Kytö, M. (2018). Heritage Swedish, English, and textual space in rural communities of practice. In J. Heegård Petersen & K. Kühl (Eds.), *Selected Proceedings of the 8th Workshop on Immigrant Languages in the Americas (WILA 8)* (pp. 44–54). Somerville, MA: Cascadilla Proceedings Project.

Hoffmann, T. (2011). *Preposition placement in English: A usage-based approach*. Cambridge: Cambridge University Press.

Huber, M. (2007). The *Old Bailey Proceedings*, 1764–1834. Evaluating and annotating a corpus of 18th- and 19th-century spoken English. In A. Meurman-Solin & A. Nurmi (Eds.), *Annotating variation and change* (Studies in Variation, Contacts and Change in English 1). Helsinki: University of Helsinki: Unit for Variation, Contacts and Change in English. See <www.helsinki.fi/varieng/series/volumes/01/huber/>

Hundt, M. & Leech, G. (2012). "Small is beautiful": On the value of standard reference corpora for observing recent grammatical change. In T. Nevalainen & E. Traugott (Eds.) *The Oxford handbook of the history of English* (pp. 175–188). Oxford: Oxford University Press.

Hundt, M., & Mair, C. (1999). 'Agile' and 'uptight' genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics*, 4, 221–242; reprinted in D. Biber & R. Reppen (Eds.), *Corpus linguistics* [Sage Benchmarks in Language and Linguistics 3: Varieties] (pp. 199–216). Los Angeles: Sage, 2012.

Jaffe, A. (2015). Book review of Barbara Johnstone, *Speaking Pittsburghese. The story of a dialect, 2013. Journal of Sociolinguistics* 19(4), 559–582.

Jucker, A. H., & Kopaczyk, J. (2013). Communities of practice as a locus of language change. In J. Kopaczyk & A. H. Jucker (Eds.), *Communities of practice in the history of English* (pp. 1–16). Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.235.01int

Kabatek, J. (2018). *Lingüística coseriana, lingüística histórica, tradiciones discursivas*. Madrid & Frankfurt: Iberoamericana & Vervuert.

Koch, P., & Oesterreicher, W. (1985). Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanisches Jahrbuch* 36, 15–43 [English translation available: Language of immediacy – language of distance: Orality and literacy from the perspective of language theory and linguistic history. In C. Lange, B. Weber, & G. Wolf (Eds.), *Communicative spaces – Variation, contact, and change: Papers in honour of Ursula Schaefer* (pp. 441–473). Frankfurt: Lang, 2012].

Kytö, M. (1991). *Variation and diachrony, with early American English in focus: Studies on CAN/MAY and SHALL/WILL* (Bamberger Beiträge zur Englischen Sprachwissenschaft 28 / University of Bamberg Studies in English Linguistics 28). Frankfurt: Peter Lang.

Kytö, M., Grund, P. J., & Walker, T. (2011). *Testifying to language and life in Early Modern England*. Including a CD-ROM containing *An Electronic Text Edition of Depositions 1560–1760 (ETED)*. Amsterdam: John Benjamins. https://doi.org/10.1075/z.162

Kytö, M., & Pahta, P. (2012). Evidence from historical corpora up to the twentieth century. In T. Nevalainen & E. C. Traugott (Eds.), *The Oxford handbook of the history of English* (pp. 123–133). Oxford: Oxford University Press.

Kytö, M., & Pahta, P. (2016). *Cambridge handbook of English historical linguistics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139600231

Kytö, M., & Smitterberg, S. (2015). Diachronic registers. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 330–345). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.019

Lange, M. B. (2009). Texts and text types in the history of German. In G. T. Horan (Ed.), *Landmarks in the history of the German language* (pp. 113–136). Oxford: Peter Lang.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press. First published in 1990 as *Institute for Research on Learning* report 90–0013. https://doi.org/10.1017/CBO9780511815355

Leech, G. (2007). New resources, or just better old ones? The holy grail of representativeness. In M. Hundt, N. Nesselhauf & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 133–149). Cambridge: Cambridge University Press. https://doi.org/10.1163/9789401203791_009

Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511642210

Leech, G., & Smith, N. (2005). Extending the possibilities of corpus-based research in the twentieth century: A prequel to LOB and FLOB. *ICAME Journal*, 29, 83–98.

Lutzky, U. (2012). *Discourse markers in Early Modern English* (Pragmatics & Beyond New Series 227). Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.227

Mair, C. (1997). Parallel corpora: A real-time approach to the study of language change in progress. In M. Ljung (Ed.), *Corpus-based studies in English: Papers from the Seventeenth International Conference on English Language and Research on Computerized Corpora (ICAME 17)* (pp. 195–209). Amsterdam: Rodopi.

Mair, C. (2016). Audio recordings. In M. Kytö & P. Pahta (Eds.), *The Cambridge handbook of English historical linguistics* (pp. 146–163). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139600231.010

Martin, P. (1990). The phrasal verb: Diachronic development in British and American English. Unpublishd PhD dissertation, Columbia University Teachers College (USA).

Marttila, V. (2014). Creating digital editions for corpus linguistics: The case of Potage Dyvers, a family of six Middle English recipe collections. Unpublished PhD dissertation, University of Helsinki. <https://helda.helsinki.fi/handle/10138/135589>

McEnery, A., & Baker, H. (2017). *Corpus linguistics and 17th-century prostitution: Computational linguistics and history*. London: Bloomsbury.

McIntosh, C. (1998). *The evolution of English prose 1700–1900: Style, politeness, and print culture*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511582790

Meurman-Solin, A. (1993). *Variation and change in early Scottish prose. Studies based on the Helsinki Corpus of Older Scots* (Annales Academiae Scientiarum Fennicae 65). Helsinki: Academiae Scientiarum Fennicae.

Meurman-Solin, A. (2007). *Manual to the Corpus of Scottish Correspondence*. Helsinki. <http://www.helsinki.fi/varieng/csc/manual/>

Meyerhoff, M. (2002). Communities of practice. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 526–548). Malden, MA: Blackwell.

Nevalainen, T., & Raumolin-Brunberg, H. (1989). A corpus of Early Modern Standard English in a socio-historical perspective. *Neuphilologische Mitteilungen*, 90(1), 67–111.

Nevalainen, T., & Raumolin-Brunberg, H. (1993). Early Modern British English. In M. Rissanen, M. Kytö & M. Palander-Collin (Eds.) *Early English in computer age: Explorations through the Helsinki Corpus* (Topics in English Linguistics 11) (pp. 53–73). Berlin: Mouton de Gruyter.

Nevalainen, T., & Raumolin-Brunberg, H. (2003). *Historical sociolinguistics: Language change in Tudor and Stuart England*. London: Longman.

Oxford English dictionary. 3rd ed. in progress: *OED* Online, March 2000–, J. A. Simpson (Ed.). <www.oed.com/>

Pelli, M. G. (1976). *Verb-particle combinations in American English: A study based on American plays from the end of the 18th century to the present*. Bern: Francke.

Rissanen, M. (1967). *The uses of one in Old and Middle English* (Mémoires de la Société Néophilologique de Helsinki 31). Helsinki: Société Néophilologique.

Rissanen, M. (1997). The pronominalization of *one*. In M. Rissanen, M. Kytö & K. Heikkonen (Eds.), *Grammaticalization at work: Studies in long-term developments in English* (pp. 87–143). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110810745.87

Rissanen, M. (2012). Corpora and the study of the history of English. In M. Kytö (Ed.), *English corpus linguistics: Crossing paths* (pp. 197–220). Amsterdam: Rodopi. https://doi.org/10.1163/9789401207935_011

Rissanen, M., Kytö, M., & Heikkonen, K. (1997). *English in transition: Corpus-based studies in linguistic variation and genre styles*. Berlin: De Gruyter. https://doi.org/10.1515/9783110811148

Rodríguez-Puente, P. (2019). *The English phrasal verb, 1650–present: History, stylistic drifts, and lexicalisation*. Cambridge: Cambridge University Press.

Romaine, S. (1982). *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511720130

Rühlemann, C., & Hilpert, M. (2017). Colloquialization in journalistic writing. *Journal of Historical Pragmatics*, 18(1), 101–135. https://doi.org/10.1075/jhp.18.1.05ruh

Rydén, M. (1966). Relative constructions in early sixteenth century English: With special reference to Sir Thomas Elyot (Studia Anglica Upsaliensia 3). Uppsala: Acta Universitatis Upsaliensis.

Schendl, H. (1997). Morphological variation and change in Early Modern English: *my/mine, thy/thine*. In R. Hickey & S. Puppel (Eds.), *Language history and linguistic modelling. A festschrift for Jacek Fisiak on his 60th birthday* (pp. 179–191). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110820751.179

Smitterberg, E. (2005). *The progressive in 19th-century English: A process of integration*. Amsterdam: Rodopi. https://doi.org/10.1163/9789004333086

Smitterberg, E. (2016). Extracting data from historical material. In M. Kytö & P. Pahta (Eds.), *The Cambridge handbook of English historical linguistics* (pp. 181–199). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139600231.012

Spasov, D. (1966). *English phrasal verbs*. Sofia: Naouka Izkoustvo.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. M. (2004). *Research genres: Exploration and applications*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524827

Taavitsainen, I. (1997). Genre conventions: Personal affect in fiction and non-fiction in Early Modern English. In M. Rissanen, M. Kytö, & K. Heikkonen (Eds.), *English in transition: Corpus-based studies in linguistic variation and genre styles* (pp. 185–266). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110811148.185

Taavitsainen, I. (2001). Changing conventions of writing: The dynamics of genres, text types, and text traditions. *European Journal of English Studies*, 5(2), 139–150. https://doi.org/10.1076/ejes.5.2.139.7309

Taavitsainen, I. (2016). Genre dynamics in the history of English. In M. Kytö, & P. Pahta (Eds.), *The Cambridge handbook of English historical linguistics* (pp. 271–285). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139600231.017

Taavitsainen, I., Kytö, M., Claridge, C., & Smith, J. (2014). English in the digital age: A general introduction. In I. Taavitsainen, M. Kytö, C. Claridge, & J. Smith (Eds.), *Developments in English: Expanding electronic evidence* (pp. 1–8). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139833882.002

Taavitsainen, I., Kytö, M., Claridge, C., & Smith, J. (2014). *Developments in English: Expanding electronic evidence*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139833882

Thim, S. (2006). Phrasal verbs in Late Middle and Early Modern English: Combinations with *back, down, forth, out* and *up*. In C. Dalton-Puffer, D. Kastovsky, N. Ritt, & H. Schendl (Eds.), *Syntax, style and grammatical norms: English from 1500–2000* (pp. 213–228). Bern: Peter Lang.

Thim, S. (2012). *Phrasal verbs. The English verb-particle construction and its history*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110257038

Todorov, T. (1990). *Genres in discourse*. Cambridge: Cambridge University Press.

Traugott, E. C. (2008). The state of English language studies: A linguistic perspective. In M. Thormählen (Ed.), *English now: Papers from the 20th IAUPE Conference in Lund* (pp. 199–225). Lund: Centre for Languages and Literature, Lund University.

von Polenz, P. (1999). *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart, Vol. 3: 19. und 20. Jahrhundert*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110805918

Walker, T. (2007). *Thou and you in Early Modern English dialogues: Trials, depositions, and drama comedy*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.158

Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics: A symposium* (pp. 95–195). Austin, TX: University of Texas Press.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511803932

Werlich, E. (1982). *A text grammar of English*, 2nd ed. Heidelberg: Quelle & Meyer.

Widlitzki, B., & Huber, M. (2016). Taboo language and swearing in 18th century and 19th century English. A diachronic study based on the Old Bailey Corpus. In M. J. López-Couso, B. Méndez-Naya, P. Núñez-Pertejo, & I. M. Palacios-Martínez (Eds.), *Corpus linguistics on the move: Exploring and understanding English through corpora* (pp. 313–336). Amsterdam: John Benjamins. https://doi.org/10.1163/9789004321342_015

Wild, K. (2010). Attitudes towards English usage in the Late Modern Period: The case of phrasal verbs. Unpublished PhD dissertation, University of Glasgow.

Wright, S. (1994). The place of genre in the corpus. In M. Kytö, M. Rissanen & S. Wright (Eds.), *Corpora across the centuries* (pp. 101–107). Amsterdam: Rodopi.

Yáñez-Bouza, N. (2015a). *Preposition stranding and prescriptivism in English from 1500 to 1900: A corpus-based approach*. Cambridge: Cambridge University Press.

Yáñez-Bouza, N. (2015b). 'Have you ever written a diary or a journal?' Diurnal prose and register variation. *Neuphilologische Mitteilungen*, 116(2), 449–474.

Address for correspondence

Merja Kytö
Department of English
Uppsala University
Box 527
SE-751 20 Uppsala
Sweden
merja.kyto@engelska.uu.se