

7 English Genres in Diachronic Corpus Linguistics

Erik Smitterberg & Merja Kytö

Uppsala University

1. Introduction

This chapter is about problems and possibilities associated with using genre as a parameter in corpus-based historical linguistics. We will begin by discussing why genre has become an increasingly central concept in historical linguistics and by defining genre and related terms. Next, we will discuss a number of challenges that corpus linguists need to address when they use genre as a parameter in their research. To begin with, we will discuss potential conflicts between two key desiderata, namely representativity and comparability. We will also take up the problem that not all genres are attested for the whole history of English and that even genres which have a long history may have changed over time. Thirdly, we will discuss how historical linguists have used genre comparisons to access an approximation of past speech. We will then devote a section to two case studies where genre plays a central role. We first look at how additional information can be gained by considering genre differences within the framework of multi-feature approaches to genre variation in the past; this account is followed by an analysis of a single linguistic feature, viz. the units co-ordinated by *and*. The chapter ends with a summary of our main points and some desiderata for future work.

2. The Centrality of the Genre Concept

Even though linguistic variation according to genre was recognized as an important variable before the advent of corpus linguistics, it has become even more central in corpus-based approaches (see Lange 2012: 401).

How to cite this book chapter:

Smitterberg, E. and Kytö, M. 2015. English Genres in Diachronic Corpus Linguistics. In: Shaw, P., Erman, B., Melchers, G. and Sundkvist, P. (eds) *From Clerks to Corpora: essays on the English language yesterday and today*. Pp. 117–133. Stockholm: Stockholm University Press. DOI: <http://dx.doi.org/10.16993/bab.g> License: CC-BY.

This is so partly because every corpus-compilation project has to take the genre parameter into account. If the researcher is compiling a single-genre corpus, delimiting the genre sampled is crucial in order to reach reliable results. And if the corpus project includes several genres, considering genres in relation to one another is a key issue when the compiler decides what research questions studies based on the corpus can hope to answer.

According to Kohnen (2001: 115), genres can serve as vehicles for spreading language change. That is, while genres themselves may not bring about change, they can certainly affect whether or not a change will spread through a language. For instance, if an incoming informal feature such as the contracted form *can't* instead of *cannot* becomes accepted in informal writing, colloquial written genres can function as “bridgeheads” from which the innovative feature can colonize other forms of writing. Hundt & Mair (1999: 236) note that a new form which arises in speech often “then spreads at differential speeds through various genres until at a very remote point it can be said to have been established in ‘the language’”. At the same time, genres may also retard a change that is spreading through a language by preserving conservative and/or fossilized usage; for instance, many legal texts preserve an obligation use of *shall* that is no longer current in everyday communication.

There is a certain amount of terminological confusion in linguistic research that considers the genre parameter. Terms such as *genre*, *register*, and *text type* are sometimes used by different linguists to mean more or less the same thing. In this chapter, we use the term “genre” to refer to categories of texts that are defined on extralinguistic or text-external grounds (the term “register” has also been used to cover such categories in previous work); in contrast, we reserve the term “text type” for categories that are defined based on their linguistic characteristics.

Within this framework, the linguistic make-up of the text itself thus does not determine what genre it belongs to; for instance, a novel may be written in the form of a series of letters or diary entries and still remain a novel. This means that genres are “fuzzy sets”: central members of a genre category will have a form that is close to the genre prototype, while more peripheral members of the set will deviate from the prototypical pattern. For instance, a prototypical member of a genre such as academic writing will contain a relatively large number of linking adverbials, prepositional phrases, and passive clauses. However, text types do not have to correlate with genres; it would be fully possible, for instance, to write a novel where the text type was Scientific

Exposition, although it would not be a prototypical member of the genre; Susanna Clarke's 2004 novel *Jonathan Strange & Mr Norrell*, for instance, remains a historical fantasy novel even though it contains numerous footnotes, a feature not usually associated with that genre. Other criteria for genre membership include the function of the texts in the society in which they are used (a novel, for example, may be read primarily for amusement) and audience expectations (members of the audience at a linguistic symposium, for instance, will typically expect a paper to inform them with regard to some aspect of linguistics). What we will mainly focus on in the remainder of this section is a set of problems that historical linguists typically encounter when including genre as a parameter in their research.

We will begin by addressing the issue of representativity. If a corpus is representative, the study of that corpus (or combination of corpora) "can stand proxy for the study of some entire language or variety of a language" (Leech 2007: 135). However, while this notion is simple enough to define, the practical application is problematic, especially with regard to historical texts (see, for instance, Biber 1993 and Leech 2007 for different suggestions on how to operationalize this parameter). Several genres – e.g. everyday conversation – are absent from the historical record even though they were important components of the language variety that researchers wish to represent, and even attested genres contain only a few of the textual witnesses that once existed.

Another desideratum of corpus-based research is comparability, viz. the extent to which sets of material are equivalent except for one single variable (Leech 2007: 141): in the case of historical linguistics, that variable is usually time. A typical example of comparability concerns the "clones" of the LOB corpus with texts from 1961; roughly equivalent publications have been sampled from later years (FLOB) as well as earlier decades (e.g. BLOB-1931) at intervals of approximately 30 years.

However, one important problem in diachronic corpus design is that representativity and comparability may clash. One reason why this happens is that genres develop and change through time, as shown in Figure 1. In Figure 1, the two ellipses represent the textual universes of newspaper English in two different periods. The difference in horizontal position indicates genre development: the textual universe in period 2 is greater, as new genres have been added (e.g. interviews), but a few genres such as shipping news have also all but disappeared during the time that separates the periods. If precedence is given to comparability,

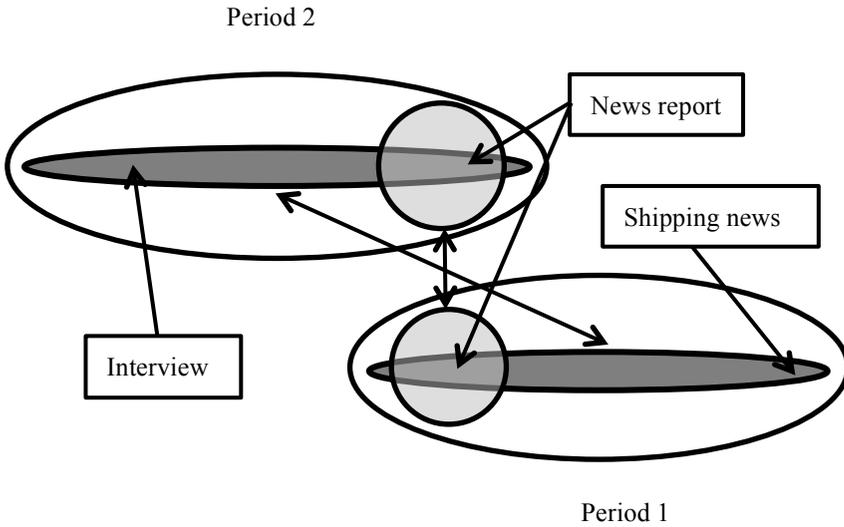


Figure 1. Textual universes of newspaper language in two periods.

only the genres that are present in both periods should be sampled (as illustrated by the two circles); this would make the period samples maximally comparable, but each sample would become less representative of the language of its period. The other strategy would be to make each period sample representative (as illustrated by the two narrow ellipses), which would instead decrease the comparability of the period samples.

Whole genres may even change across time to serve language users' needs. Drama texts included in the Helsinki Corpus (1420–1500) may serve as an example. In the Late Middle English section of the corpus, the Drama texts included are religious mystery plays, while the Early Modern English section contains Drama comedies (this difference between the samples is also recognized by the corpus compilers in that “Drama” has been qualified with different subtitles in these two cases). Both mystery plays and comedies are Drama texts in the sense that they are scripted dialogue texts used for stage performances. However, differences in the intended main functions of these two forms of drama (religious instruction vs. entertainment) may decrease the comparability of such samples from a corpus-linguistic standpoint. On the other hand, including such heterogeneous genres will make the corpus more representative of late Middle and early Modern English.

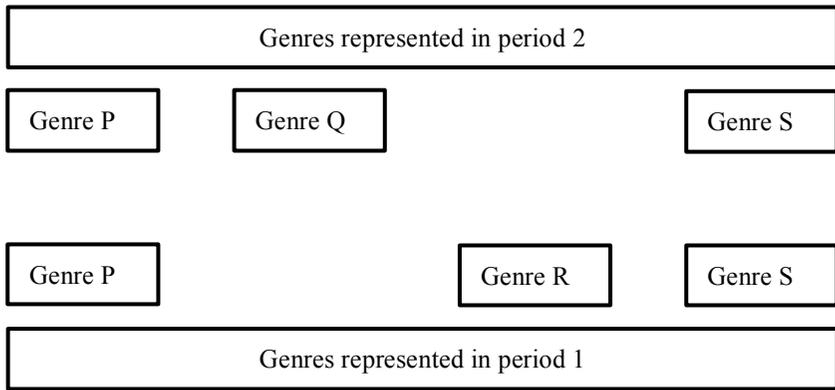


Figure 2. Variation in genre representation between two periods.

In addition, even if genres do not change over time, they may emerge late in the period studied or die out, leading to other representativity and comparability issues. This is illustrated in Figure 2. If the aim is to produce two comparable period corpora, only genres P and S should be included in corpus compilation, as they are represented in both periods. However, this would lead to each period sample being less representative, as genre Q in period 2 and genre R in period 1 would be ignored.

As Figure 2 implies, not all genres have been attested throughout the recorded history of English; for instance, the emergence of some genres is tied to technological developments (e.g. e-mail). The patchy picture that some genres afford researchers is illustrated in Figure 3.

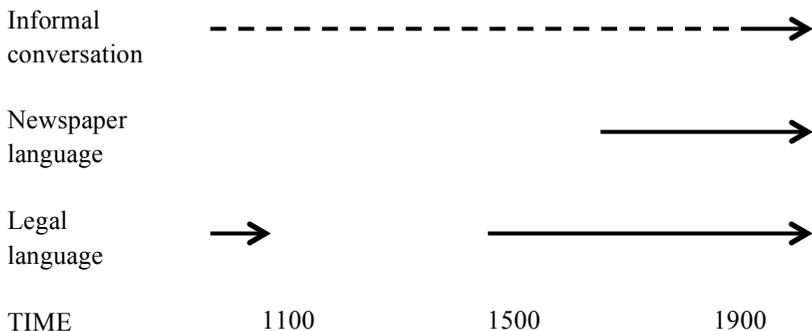


Figure 3. Genre representation in English across time.

Figure 3 illustrates one genre and two genre groups. While informal conversation has of course existed since the beginnings of the English language, it is only recorded from the 20th century on; the line is thus dashed for most of the history of the language. Genres that belong to newspaper language can be said to have existed in English since 1665, when the *Oxford Gazette* was first published, although there were precursors of newspapers before then, such as corantos (see Dahl 1953 and Brownlees 2012 for accounts of the beginnings of periodical news). Finally, genres may disappear at a certain point in the history of a given language only to re-emerge later on. This is the case with law texts in English. They are attested in the Old English period, but are absent from most of the Middle English period, as Latin and, later, French took over in official use in England. However, after a gap of several centuries, law texts written in English re-appeared in the late Middle English period (Claridge 2012: 239–240).

Yet another scenario that may introduce problems for corpus compilers is one in which two or more genres exist in all periods studied, but their relative importance has changed over time. The question then becomes whether this change should be reflected in corpus compilation. Figure 4 illustrates this situation. The news report did exist in seventeenth-century England: as mentioned above, the *Oxford Gazette*, later the *London Gazette*, appeared in 1665, and there were precursors such as corantos. But it was not a central genre compared with the Bible,

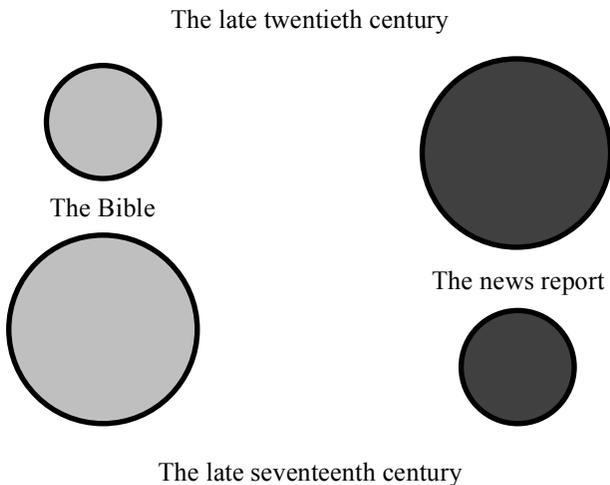


Figure 4. The relative importance of two genres in late seventeenth-century and late twentieth-century English.

which was part of most language users' daily lives. The situation had arguably been reversed by the late twentieth century: the secularized nature of modern Britain means that most speakers have limited contact with biblical texts, while news in some form – newspapers, television, the Internet, etc. – reaches nearly all language users. In addition, multiculturalism has led to the establishment of religions that do not make use of the Bible in Britain, which further decreases the special standing of that text.

There are thus clear difficulties involved both in achieving representativity and in balancing comparability and representativity. Depending on the underlying research questions that the study of a given corpus is intended to help answer, corpus compilers have approached these difficulties from different angles. For instance, the compilers of the BLOB-1931 clone of the LOB corpus gave precedence to comparability over representativity in the selection of newspaper texts (Leech 2007: 143): priority was given to including the same newspapers in BLOB that had been included in LOB, even though a given newspaper may not have been equally representative of its genre around 1931 and in 1961. However, as Leech (2007: 143) notes, in investigations with greater time-depth, complications relating to genre evolution would have made comparability very difficult to achieve. In such cases, varying the granularity of the genre parameter may be useful. When possible, corpus texts may be coded not only for genre, but also for subgenre and/or subfunction. Biber and Gray (2013) argue that keeping subgenres such as newspaper articles and news magazine articles constant may be essential in order to allow researchers to identify and describe language-change phenomena with a high degree of reliability. As regards subfunctions, Kohonen (2007) suggests that coding parts of corpus texts according to the subfunction they fill can help to make texts more comparable; for instance, a genre like religious instruction can be divided into subfunctions such as narration and exegesis, and texts belonging to the same genre can then be compared to see whether they also emphasize the same subfunctions.

Alternatively, grouping several genres together into hyperonymic entities can enable scholars to collect roughly comparable text categories if not all genres are attested in all periods sampled. The prototypical text categories in the Helsinki Corpus can be used as an example of this; for example, the category Imaginative Narration covers genres such as Fiction, Romance, and Travelogue, which suggests that these genres share some features that may make them roughly comparable if one or several of them are missing from some period samples.

In contrast, if priority is given to representativity, some steps have been taken to improve a corpus in this regard. These include covering a wide range of genres, giving precedence to texts that are considered good representatives of their genres, aiming at a proportional representation of the genres included, and simply enlarging the size of the corpus (see Kytö & Smitterberg forthcoming for a fuller discussion). In historical linguistics, the most serious obstacle to achieving representativity is of course the lack of spoken texts. The available material has been preserved in writing, while a great deal of the actuation of change is likely to have taken place in speech and notably in speech used in dialogue situations (see, however, Biber & Gray 2011 for an account of change that has spread mainly in non-speech-related writing). One solution has been to turn to comparisons of texts and genres that stand at different distances from past speech. For instance, written records of spoken language can be assumed to come closer to the actual spoken language of the time than written language that was not based on language taken down in speech situations or created to imitate speech. Even though it is practically impossible to take down speech in writing in all its aspects, previous research indicates that the essence of what was said was relayed in the recorded version. Consequently, it is necessary to look for what have been referred to as “spoken”, “oral” or “colloquial” genres. To this end, scholars have compiled genre-specific corpora that focus on speech-related language. For example, in the Corpus of English Dialogues 1560–1760 (CED), past speech can be studied from the perspective of speech-purposed (e.g. Drama) and speech-based (e.g. Depositions) genres (see Culpeper & Kytö 2010 for more detailed information on CED, including case studies). The results of such analyses can then be compared with those based on speech-like genres such as Private Correspondence in order to shed further light on the spoken language of the period.

3. Case Studies: Drama and Science in Focus

We will now consider two actual studies to show how the impact of the genre perspective has affected historical corpus linguistics. Our examples are taken from the Modern English period. One study considers the co-variation of a large number of features, while the other focuses on a single linguistic feature.

One way of looking at linguistic variation is to consider how a large number of linguistic features co-vary in texts. Within this framework,

different features load on a number of dimensions of variation. Features that co-occur in texts end up on the same pole on a dimension of variation; features that tend not to co-occur also belong to the same dimension, but will be placed on opposite ends. When genres are positioned on these dimensions, they can be shown to be more or less “involved”, “informational”, etc., depending on the co-occurrence patterns of the linguistic features that are included in the analysis. Among the pioneers in extending this methodology to historical texts are Douglas Biber and Edward Finegan; we will consider one of their studies here, viz. Biber & Finegan (1997). This study is based on the original version of A Representative Corpus of Historical English Registers (ARCHER), a 1.7-million-word corpus of British and American English which covers the period 1650–1990. The features studied include colloquial features such as contractions as well as features characteristic of impersonal styles, e.g. passives (Biber & Finegan 1997: 258–259).

In their study, Biber and Finegan (1997) demonstrate that not all English genres have followed the same trajectory through Modern English. Whereas specialist expository writing has consistently tended towards more “literate” styles, popular non-expository texts show a reversal of this trend towards more “oral” styles during the Late Modern English period (Biber & Finegan 1997: 272–273). As a result, the linguistic differentiation between different kinds of writing increases during the period studied. For instance, while Drama leads the way towards a more involved style of communication, science writing exhibits an equally clear trend towards the informational end of the same dimension (Dimension 1, “Involved versus Informational Production”) (Biber & Finegan 1997: 266).

Another way of considering linguistic variation is from the perspective of a single linguistic feature. The second case study, which concerns the co-ordinator *and* and the linguistic units that *and* can co-ordinate, exemplifies this perspective. Following Quirk et al. (1985), we will refer to these units as “conjoints”. *And* can link conjoints on different levels of syntactic structure, from individual morphemes to whole sentences. In the literature, a basic distinction is often made between clausal co-ordination and phrasal co-ordination (see, for instance, Culpeper & Kytö 2010). In example (1) below, the co-ordination is clausal because the two conjoints are main clauses. In example (3), in contrast, the two conjoints are adjective phrases, so we are dealing with phrasal co-ordination. We apply a similar, but slightly modified classification to the data in this study.

The main reason why co-ordination by *and* is of interest is that “oral” and “literate” genres display different preferences as regards phrasal and clausal co-ordination. Biber et al. (1999) show that, while clausal co-ordination predominates in conversation in Present-day English, phrasal co-ordination is frequent in academic writing. Moreover, Biber (2003) demonstrates that clausal co-ordination and phrasal co-ordination are characteristic of “oral” and “literate” discourse, respectively. We might thus expect expository genres to display more phrasal co-ordination than non-expository genres in Late Modern English.

To investigate this, Smitterberg (forthcoming) retrieved a random subset of *and* in A Corpus of Nineteenth-century English (CONCE), a one-million-word multi-genre corpus of British English from the 1800s. We focus on results for Drama and Science, one stereotypically “oral” and one stereotypically “literate” genre. The analysis is based on 400 randomly selected instances per genre from each of the two periods included, viz. 1800–1830 and 1870–1900. The conjoins of each instance of *and* retrieved were classified according to their syntactic make-up. Three categories are recognized in this classification. Smitterberg refers to the first category, which is exemplified in (1) and (2), as “super-phrasal”. (Conjoins are given in bold face in numbered examples; speaker identifications and stage directions are enclosed in square brackets and dollar signs in the corpus.)

- (1) [\$Blunt.\$] No. **I came too late, and I am sorry for it:** [...] (CONCE, Drama, Holcroft, 1800–1830, p. 25)
- (2) Soon after the application of the heat, a dark line, thin and delicate as a spider’s thread, was observed to be **slowly creeping down each of the bright sodium lines and exactly occupying the centre of each.** (CONCE, Science, Lockyer, 1870–1900, p. 128)

In order to be included in this category, the conjoins have to meet two criteria. First, both have to contain more material than one syntactic phrase. Second, both conjoins have to contain at least part of a verb phrase. For the second category, the traditional term “phrasal” is used; it is illustrated in (3) and (4).

- (3) [\$MISS T.\$] Have those **base and servile** things called settlements been satisfactorily adjusted? [\$eating\$] (CONCE, Drama, Gilbert, 1870–1900, p. 25)
- (4) It would undoubtedly be advantageous to the capitalists of England, and to the consumers in both countries, that under

such circumstances, **the wine and the cloth** should both be made in Portugal, [...]

(CONCE, Science, Ricardo, 1800–1830, pp. 160–161)

In phrasal co-ordination, the conjoins are either on or below the level of a syntactic phrase and do not consist of full verb phrases. Finally, Smitterberg (forthcoming) recognizes an indeterminate category, to which examples (5) and (6) belong. This category contains examples whose conjoins did not meet all criteria for either of the other categories.

- (5) [\$Mait.\$] A vindictive temper is the master passion that **degrades and ruins** the peace of Mr. Anson: [...]

(CONCE, Drama, Holcroft, 1800–1830, p. 34)

- (6) [\$Admiral. [More sandwich.]]\$ If ever there was a jewel of a wife it's Lady Darby. God bless her! Here's her health. [\$Drinks.\$] I don't deserve her. She's too good for me. When I remember **what an unfaithful rascal I've been, and the lies I've had to tell** – the awful lies – [\$Is overcome with painful reminiscences and weeps.\$]

(CONCE, Drama, Jones, 1870–1900, p. 50)

In (5), both conjoins contain one verb phrase but no other material; and in (6), the first conjoin consists of a clause and the second of a noun phrase.

The manual analysis of 400 instances from each period/genre sample led to the exclusion of 116 instances, or c. 7% of the data. These include instances of *and* in stage directions, chapter headings, and numerical expressions such as *four and a half*; a small number of

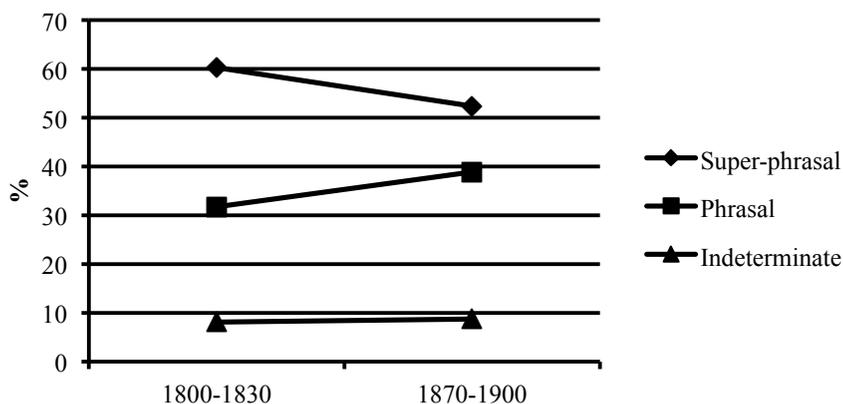


Figure 5. Conjoins of *and* in Drama.

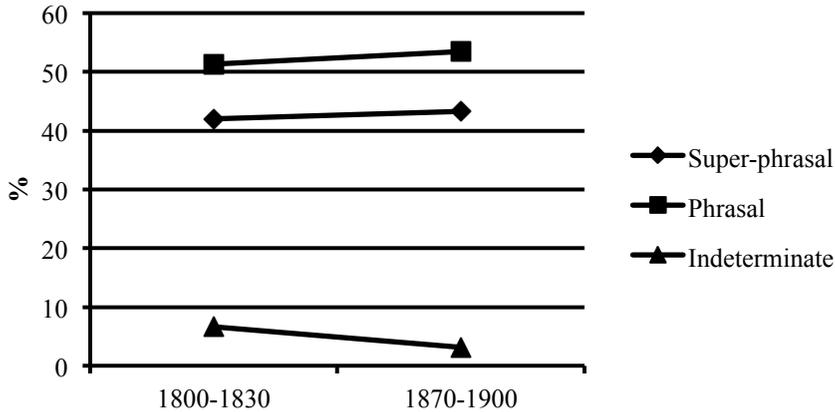


Figure 6. Conjoins of *and* in Science.

instances that resisted classification were also excluded. The total number of instances of *and* included is 347 from Drama, period 1, 378 from Drama, period 3, 376 from Science, period 1, and 383 from Science, period 3. The period/genre distribution is given in Figures 5 and 6.

As Figures 5 and 6 show, the proportion of super-phrasal co-ordination is higher in Drama than in Science; this difference is statistically significant in both periods (for period 1, d.f. = 2; $\chi^2 = 28.9$; $p < 0.001$; for period 3, d.f. = 2; $\chi^2 = 22.1$; $p < 0.001$). This genre difference tallies well with previous research on spoken and written communication. As mentioned above, Biber's (2003) factor analysis of present-day academic English demonstrates that clausal co-ordination and phrasal co-ordination are characteristic of spoken and written English, respectively. As Drama comes out as a stereotypically "oral" genre in Biber and Finegan's (1997) diachronic factor score analysis of Modern English, the predominance of super-phrasal co-ordination in this genre is to be expected. Likewise, the high percentage of phrasal co-ordination in the "literate" Science genre is in accordance with what would be expected against the background of Biber's (2003) results.

In contrast to the cross-genre differences, there are no clear indications of change across time in either genre. While there are tendencies towards, for instance, more phrasal co-ordination in Drama and less indeterminate co-ordination in Science, these period differences do not reach statistical significance (for Drama, d.f. = 2; $\chi^2 = 4.72$; $p = 0.095$; for Science, d.f. = 2; $\chi^2 = 5.06$; $p = 0.080$). Neither the consistent trend towards more literate styles in specialist expository writing nor the

reversal towards more oral styles in a popular written genre such as Drama noted by Biber and Finegan (1997: 272–273) is thus mirrored in our results. Regarding the results for Science, Geisler (2002) found that this genre did not change significantly on any dimension in his factor score analysis of CONCE; it is possible that differences in textual selection and/or the time span covered underlie these differences between results found for scientific English in CONCE and ARCHER (cf. Biber & Finegan 1997).

The genre perspective comes across as crucial in two respects in our case study on co-ordination. First, including both “oral” and “literate” writing is necessary to obtain a full picture of co-ordination in nineteenth-century English as a whole; neither genre included could have stood proxy for the entire language variety. Secondly, it is well known that, while language change presupposes language variation, not all variation leads to change; a genre perspective can help to uncover such cases of stable variation in language.

The two approaches exemplified here both yield important information on the development of the English language and on how this development is connected to the genre parameter. For instance, detailed studies of single features can show what features are worth including in multi-feature analyses; at the same time, the overall view afforded by multi-dimensional studies provide single-feature analyses with an overall theoretical framework (see Biber 1988: 62–63 for further discussion).

4. Concluding Remarks

As we hope to have shown in this chapter, genre is an indispensable parameter in historical linguistics. If the language of a period is treated as a monolithic phenomenon, patterns such as genre drift and genre differentiation may go unnoticed. Similarly, factors underlying the distribution of individual linguistic features may escape notice if genre differences in their occurrence are not taken into account.

However, as we have also demonstrated, the limitations imposed by historical material require that attention be paid to methodological issues. Above all, it is crucial that researchers account for their definitions of the genre concept and their criteria for identifying and classifying genres (e.g. the relative importance of linguistic and extralinguistic criteria). Attention must also be paid to the socio-historical contexts of the genres sampled; for instance, factors such as literacy and level of

education will have affected the size of the audience of a written genre at any point in the history of English. Conversely, the author perspective is important especially when dealing with genres that have few textual witnesses. A text such as the *Ormmulum*, for instance, can be seen as representing the genre of homily collections in the twelfth century. However, it can also be seen as a concrete representation of the author's own idiolect as against the more abstract genre level.

As we have discussed above, the desiderata of comparability and representativity may clash when corpora are compiled from a genre perspective. Genres have emerged and died out through the recorded history of English, causing gaps in genre representation across time. There are of course also plenty of examples of genre continuity; but even in such cases, attention must be paid to genre evolution and the extent to which what is nominally the "same" genre can be said to occupy the same communicative space in an ever-changing society.

The two case studies we have reported on demonstrate that the genre concept is of central importance in diachronic corpus linguistics. This is of course true of multi-feature studies, which are typically based on contrasting linguistic co-occurrence patterns in several different genres. However, even studies of single linguistic features such as *and* and its conjoins often require a genre perspective, as different patterns may manifest themselves in different genres and genre groups.

As regards future developments, the genre concept is likely to continue to grow in significance in historical corpus linguistics. Above all, genres are likely to become central parameters in a wide range of sub-disciplines, from pragmatics and discourse studies to grammar. (See, for instance, Walker 2007 for a study where the genre concept is crucial for an understanding of pragmatic variation and change in pronoun usage.) Regarding methodology, there is a need for new approaches to studying past language forms on their own terms. In terms of resources, there is plenty to do in, for instance, providing faithful linguistic editions of early manuscripts that can be used as the basis for new historical corpora. We also need to pay better attention to poorly represented speaker groups, for instance the language of women and lower socio-economic strata. In addition to studying British English, we should also consider the development of genres in overseas varieties of the language.

Corpora Referred to

ARCHER = A Representative Corpus of Historical English Registers, originally compiled under the supervision of D. Biber and E. Finegan (modified and expanded by subsequent members of a consortium of universities). For more details, see <http://www.alc.manchester.ac.uk/subjects/lcl/research/projects/archer/> and <http://www.helsinki.fi/varieng/CoRD/corpora/ARCHER/updated%20version/introduction.html>.

BLOB-1931 = The BLOB-1931 Corpus, compiled by G. Leech, P. Rayson & N. Smith. For more details, see <http://www.helsinki.fi/varieng/CoRD/corpora/BLOB-1931/index.html>.

CED = A Corpus of English Dialogues 1560–1760 (2006), compiled by M. Kytö & J. Culpeper. For more details, see <http://www.helsinki.fi/varieng/CoRD/corpora/CED/index.html>.

CONCE = A Corpus of Nineteenth-century English, compiled by M. Kytö & J. Rudanko. For more details, see Smitterberg (2005).

FLOB = The Freiburg–LOB Corpus of British English (1999), compiled by C. Mair. For more details, see <http://www.helsinki.fi/varieng/CoRD/corpora/FLOB/index.html>.

Helsinki Corpus = The Helsinki Corpus of English Texts (1991), compiled by M. Rissanen (Project leader) & M. Kytö (Project secretary); L. Kahlas-Tarkka & M. Kilpiö (Old English); S. Nevanlinna & I. Taavitsainen (Middle English); T. Nevalainen & H. Raumolin-Brunberg (Early Modern English). For more details, see <http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/index.html>.

LOB = The Lancaster-Oslo/Bergen Corpus (1976), compiled by G. Leech, S. Johansson & K. Hofland. For more details, see <http://www.helsinki.fi/varieng/CoRD/corpora/LOB/index.html>.

References

- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- . (1993). Representativeness in corpus design. *Literary and Linguistic Computing* 8:4, 243–257.
- . (2003). Variation among university spoken and written registers: A new multi-dimensional analysis. P. Leistyna & C. Meyer (eds) *Corpus Analysis: Language Structure and Language Use*. Amsterdam & New York: Rodopi, 47–70.

- Biber, D. & Finegan, E. (1997). Diachronic relations among speech-based and written registers in English. T. Nevalainen & L. Kahlas-Tarkka (eds) *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*. Helsinki: Société Néophilologique, 253–275.
- 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.
- . (2013). Being specific about historical change: The influence of sub-register. *Journal of English Linguistics* 41:2, 104–134.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson.
- Brownlees, N. (2012). The beginnings of periodical news (1620–1665). R. Facchinetti, N. Brownlees, B. Bös & U. Fries *News as Changing Texts: Corpora, Methodologies and Analysis*. Newcastle upon Tyne: Cambridge Scholars Publishing, 5–48.
- Claridge, G. (2012). Linguistic levels: Styles, registers, genres, text types. A. Bergs & L.J. Brinton (eds) *English Historical Linguistics: An International Handbook*. Vol. 1. Berlin & Boston: Walter de Gruyter, 237–253.
- Clarke, S. (2004). *Jonathan Strange & Mr Norrell*. New York & London: Bloomsbury.
- Culpeper, J. & Kytö, M. (2010). *Early Modern English Dialogues. Spoken Interaction as Writing*. Cambridge: Cambridge University Press.
- Dahl, F. (1953). *A Bibliography of English Corantos and Periodical Newsbooks 1620–1642*. Stockholm: Almqvist & Wiksell.
- Geisler, G. (2002). Investigating register variation in nineteenth-century English: A multi-dimensional comparison. R. Reppen, S.M. Fitzmaurice & D. Biber (eds) *Using Corpora to Explore Linguistic Variation*. Amsterdam & Philadelphia: John Benjamins, 249–271.
- Hundt, M. & Mair, C. (1999). “Agile” and “uptight” genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4:2, 221–242.
- Kohnen, T. (2001). Text types as catalysts for language change: The example of the adverbial first participle construction. H-J. Diller & M. Görlach (eds) *Towards a History of English as a History of Genres*. Heidelberg: Universitätsverlag C. Winter, 111–124.
- . (2007). From Helsinki through the centuries: The design and development of English diachronic corpora. P. Pahta, I. Taavitsainen, T. Nevalainen & J. Tyrkkö (eds) *Towards Multimedia in Corpus Studies*. Helsinki: Research Unit for Variation, Contacts and Change in English (VARIENG), University

- of Helsinki. Available at <http://www.helsinki.fi/varieng/journal/volumes/02/kohnen/>; last accessed 13 February 2013.
- Kytö, M. & Smitterberg, E. (Forthcoming). Diachronic registers. D. Biber & R. Reppen (eds) *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.
- Lange, C. (2012). Text types, language change, and historical corpus linguistics. C. Lange, B. Weber & G. Wolf (eds) *Communicative Spaces: Variation, Contact, and Change. Papers in Honour of Ursula Schaefer*. Frankfurt am Main etc.: Peter Lang, 401–416.
- Leech, G. (2007). New resources, or just better old ones? The holy grail of representativeness. M. Hundt, N. Nesselhauf & C. Biewer (eds) *Corpus Linguistics and the Web*. Amsterdam: Rodopi, 133–149.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London & New York: Longman.
- Smitterberg, E. (2005). *The Progressive in 19th-century English: A Process of Integration*. Amsterdam & New York: Rodopi.
- . (Forthcoming). *Colloquialization in Nineteenth-century English*.
- Walker, T. (2007). *Thou and You in Early Modern English Dialogues: Trials, Depositions, and Drama Comedy*. Amsterdam & Philadelphia: John Benjamins.