

A review article

Beyond the great tradition of English grammars

Nelleke Oostdijk
University of Nijmegen

Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Educated Limited, 1999. xxviii + 1204 pp. ISBN 0-582-237254. Reviewed by **Nelleke Oostdijk**, University of Nijmegen.

Introduction

Since its publication in 1985, *A Comprehensive Grammar of the English Language (CGEL; Quirk et al 1985)* has established itself as the standard reference work in the field of English grammatical description. *CGEL* is the last in a series of grammars that were produced by a team of eminent linguists, informally also known as ‘the gang of four’ – Randolph Quirk, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik, who started their collaborative work as early as the 1960s and saw the publication of *CGEL* as the culmination of their joint work. The grammar has exerted a powerful influence in the field both through its descriptive framework, its concepts and terminology, and its grammatical description.

CGEL has also had an influential role in the recently published *Longman Grammar of Spoken and Written English (LGSWE)*. As the authors acknowledge their indebtedness to *CGEL*, they also make it clear that they do not seek to compete with it: ‘While advances in corpus technology have allowed us to go beyond *CGEL* in important ways – particularly in the exemplification and quantitative investigation of grammar across different language varieties, spoken and written – *CGEL*’s attention to detail and comprehensive coverage is something to which this grammar does not attempt to aspire. In many ways, the two grammars complement rather than compete with each other’ (p viii).

The *LGSWE* is the result of a six-year project carried out by an international team of linguists who hardly need any introduction. Under the leadership of Douglas Biber, they have undertaken to write a corpus-based grammar that gives 'a thorough description of English grammar, which is illustrated throughout with real corpus examples, and which gives equal attention to the ways speakers and writers actually use these linguistic resources' (p 45). Thus the potential use of the grammar they have produced, they claim, extends beyond what was feasible with most traditional grammars. It is expected that the *LGSWE* will prove an important resource not only for descriptive studies of English, but for all sorts of studies in a much wider range of linguistic subdisciplines. More specifically, the authors refer to the grammar's potential as a resource for language teaching and its use in the field of natural language processing.

The book contains 14 chapters, organized in five sections. The first of these, Section A 'Introductory', consists of just one chapter, viz Chapter 1 entitled 'Introduction: a corpus-based approach to English grammar', in which the goals and principles of this corpus-based grammar are described quite elaborately. Sections B through D concentrate on the core grammatical description of English and the distribution of grammatical features, while Section E, appropriately entitled 'Grammar in a wider perspective', focuses on a number of selected topics that relate to the use of (alternative) grammatical structures, the grammatical marking of stance, the use of lexical expressions in speech and writing, and the interface between grammar and discourse, lexis, and pragmatics. The prominence that throughout each chapter (with the exception of Chapter 1) is given to the corpus findings and their discussion are most definitely a major asset.

Every effort has been made to make the book optimally accessible. It contains two tables of contents: the first of these, the 'contents summary' gives an overview of how the contents are structured, while the second, the 'contents in detail' lists the contents in full. In addition, there are two indexes: a lexical index and a conceptual index. While the latter provides easy access to all key terms and notions, the lexical index is useful for anyone interested in (a) the classification of specific lexical items, (b) the quantitative data provided for the occurrence of certain words, particularly in statements like 'words that occur over x times in x million words' and (c) the possible co-occurrence of one lexical item with others. Finally, I should mention the bibliography. Interestingly, it has three parts: A. 'Corpus-based studies of present-day English: general'; B. 'Corpus-informed grammars of present-day English'; and C. 'Corpus-informed studies of specific areas of present-day English grammar'. The final part is of particular

interest since it includes a great many references to articles and monographs that have appeared since the publication of *CGEL* in 1985.

Discussion

The publication of the *Longman Grammar of Spoken and Written English* (*LGSWE*) marks a point in time in the history of English language studies at which the effects of a changed linguistic setting can no longer be ignored. Since the reference grammars of the Great Tradition came into being, the perception of what constitutes the object of description, viz the description of (English) language use, has changed. The descriptions that we find in earlier grammars generally are biased towards the 'standard language'. More recently, however, the idea that English language use can be represented by a description based on a single variety was defeated in the light of corpus linguistic evidence. The variation encountered in the many and vast resources of language data available to linguists today has made it apparent that the description of language use is necessarily the description of language variation.

The shift in focus on what constitutes the object of linguistic descriptions is not without consequences. It impels us to review the adequacy of our descriptive model and descriptive apparatus, and raises a number of rather interesting methodological questions. (Corpus)linguists subscribe to the view that a qualitative description should 'not only be concerned with the study of the rule system underlying the constructions found in corpus data, but should also include, ideally, what was formulated by Quirk in 1960 (cf 1968, Chapter 7) as the 'total accountability' of texts, that is, the examination of *all* the linguistic features of a text as well as the way these features interrelate in context' (Aarts 1999: 5). And while there seems to be little dispute about the fact that, in addition to this qualitative dimension, there is also a quantitative dimension to the description of language use and variation, linguists disagree on the way in which information about the frequency and distribution of linguistic units should be incorporated. On the one hand, there is for example Leech who takes the view that 'quantities can be added (as a separate stage of description) on to any model of language, without sacrificing any of the existing features of that model' (1992: 110), and 'the quantitative values of the model are derived, after applying the model in the analysis of the corpus, from the frequency data inherent in the corpus itself' (*ibid*: 11). On the other hand, Aarts (1999) – while discussing the requirements to be met by a model for the description of language use – advocates an integral approach in which the quantitative and qualitative descriptive dimensions interact.

Against the background outlined above, it is all the more fitting that the first, introductory chapter to the *LGSWE* is devoted entirely to the corpus-based approach adopted in the book. The chapter describes the goals pursued in this grammar, while also various methodological aspects that characterize the approach are discussed and put into perspective. Because it would be an impossible task and also because I would rather focus on the corpus-based nature of the *LGSWE*, I shall not go into any detail as far as the description of specific grammatical features is concerned. Nor shall I summarize and discuss the *LGSWE* chapter by chapter. Instead, I shall organize what follows around the main issues addressed in the introduction.

In their definition of what the grammar describes, the authors have set their ambitions high: ‘The *LGSWE* describes not only the range of grammatical features in English, but also the actual use of each major feature. We consider the ways in which a feature is used, the extent of its use, and its variability in relation to other features. We also consider the factors that favor or disfavor each variant’ (p 5). The grammar focuses in particular on language use in four major registers: conversation, fiction, newspaper language and academic prose. The authors motivate their choice as follows: in answer to the question why the focus is on register variation rather than dialect variation, they argue that ‘grammatical differences across registers are more extensive than across dialects. When speakers switch between registers they are doing different things with language, using language for different purposes, and producing language under different circumstances. As a result, there are often extensive linguistic differences among registers’ (p 21). The reason for investigating these particular four registers is that they are highly productive and serve well to chart the range of variation in language use.

The grammatical descriptions in *LGSWE* are based on over 40 million words of data contained in the Longman Spoken and Written English (LSWE) Corpus. The core of the Corpus is formed by four, roughly five-million-word subcorpora, each representing one of the four main registers. For dialect comparisons, two further subcorpora are included in the Corpus, while there is also a substantial amount of data from non-conversational and general prose. The design of the corpus requires some comments, I think. Despite the elaborate discussion that is devoted to the composition of the corpus and the design criteria that were applied, a number of points remain unclear. For example, in the presentation of the Corpus we see that two of the four subcorpora (fiction and academic prose) in the core comprise both British and American data, while for the other two (conversation and news) only British English data are included. American data for the latter two registers are kept separate, apparently without good reason.

While this may be just a matter of how things are presented, other points are more puzzling. The Corpus was compiled specially for this project, in which the authors aim to describe contemporary British and American English as used by adult speakers. If ‘the present grammar does not attempt an overall treatment of dialect differences in English’ and ‘such an investigation is beyond the scope of the present grammar’ (p 26), then why does the corpus include data originating from other national varieties (incl. Australian, Caribbean and West African English)? The data are relatively few – 537,700 words of a total of 4,980,000 words – and are perhaps unlikely to skew the findings, but from a methodological point of view ‘preferably’ these data should not have been included at all. A similar observation can be made for the texts aimed at teenagers and older children that were included (450,200 words of a total of 4,529,800 words), or the speech produced by non-adults (ie young children and teenagers). Finally, while the Corpus includes mostly texts produced after 1980, the fiction texts are generally older; some of these (27 texts out of 139) appeared well before 1950. Here it would have been useful to refer to the year of publication of the editions that were actually used, rather than the year in which a book was first published. Instead, the explanation given is of an awkwardly mixed nature: while it is explained that ‘these older texts have a role in defining the receptive grammatical usage up to present day’, it is hard to see how this falls in with the observation that ‘it should be emphasized that such texts make up less than one per cent of the total number of texts’ (p 30).

Throughout the book, ample examples are given. All examples derive from the Corpus. With each example it is indicated from which register it originates and whether the example is given in full or has been truncated. However, no source references are given. While, in principle, the full Corpus comprises over 40 million words, only occasionally are all data used for investigating a particular feature. The quantitative findings reported on in the *LGSWE* are more commonly based on subsets. The ‘Endnotes’ section in the book provides detailed information on what data the analyses were based on. Unfortunately, very little information is given about how the data were processed. Quantitative data are reported as frequency counts (absolute or relative) and mutual information scores. Rather than reporting the exact number of occurrences observed, absolute frequencies are normalized to a common basis, viz occurrences per million words of text. This, according to the authors, is appropriate since there are several factors that ‘can cause minor fluctuations in the frequency counts of lexicogrammatical features Because of the influence of such factors, we report frequencies at a level of precision that we judge replicable’ (p 39). Finally, on the point of quantitative findings as reported in the grammar, it is worth noting that,

while the authors point out that the focus is ‘only on those patterns that are clearly important in addition to being statistically significant’ (p 40), there is a strong bias towards low level (linguistic) units. Quantitative findings involving tokens, types, lexical items, word classes, etc are abundant, while there is far less information for example about the frequency and distribution of (immediate constituents of) phrases. The reason for this may be that the LSWE Corpus was tagged and not, or at least not fully, parsed. A quantitative analysis of the frequency and distribution of a syntactic construct is then of course far more problematic, especially when it is difficult to identify the construct on the basis of a (combination of) particular lexical feature(s). This, however, cannot serve as an excuse for the practice adopted, for example, in reporting on the complexity of noun phrase premodification. Personally, I find it extremely awkward that complexity here is expressed in terms of the number of words instead of the number of premodifiers and/or in terms of the realization of the premodifier(s).

The *LGSWE* closely follows the descriptive framework and terminology of *CGEL*. The authors of the *LGSWE* – wanting ‘to avoid allocating too much space to justifying a descriptive framework’ – ‘have relied on previous descriptions of English’ and have settled for adopting a framework that was readily available, is ‘terminologically conservative, generally following informed tradition in its choice of grammatical terms and categories’ (pp 6–7). There are very few departures from *CGEL*, none of which can be considered ‘groundbreaking’, to put it in terms of the cover blurb. Apparently, the authors have not felt the need to critically review the descriptive framework used in *CGEL*, nor have they aspired to be innovative in this respect. While in the interpretation of the corpus findings, the authors frequently refer to discourse structures and processes, these analyses appear to be independent of any model, while there is little attempt to integrate the discourse information with the syntactic framework by adapting or extending it. Thus it can be said that the grammatical description presented in *LGSWE* seldom surpasses that in *CGEL*. In adopting the descriptive model of *CGEL*, also many of the weaknesses of *CGEL* are taken over. Moreover, we find that the *LGSWE* is not always consistent in its use and application of various descriptive notions. Especially the difference between form and function is problematic. Consider for example the following description: ‘Pronouns and determiners are closely related: they overlap in form and are both connected with the specification of reference. Often there are alternative forms, using either a determiner + noun or a pronoun. For example, many quantifiers can be used both as determiners and as pronouns, some also as adverbs’ (p 71). While the introduction to phrase grammar and clause grammar in Chapters 3/4 respectively, on the whole provides an adequate descriptive and terminological basis,

in subsequent chapters there does not appear to be a practice of systematically associating with each constituent a function and a category label. The use of function and category label is often confused and there is a strong tendency to omit the phrase level.

Earlier, I mentioned the fact that the *LGSWE* includes a great many examples. It should be noted, however, that very few instances are actually analysed in detail. This is all the more disturbing, since for some structures particular aspects of the description remain implicit, while there are also certain aspects of structures that the grammar does not cater for at all. Thus we can only guess at the analysis of the noun phrase postmodifying clause in instances such as ‘*There’s so many things that I know [I want to learn <->]*’ or ‘*That’s the bit that we don’t tend [to know so much about <->]*’ (p 623). These are instances of what the *LGSWE* describes as relative clause constructions with deeply embedded gap positions (in the examples <-> marks the gap, while the [] marks the clause). These and other corpus examples that are presented serve well to illustrate a large variety of interesting and sometimes quite complex constructions that occur in authentic language use.¹ However, upon attempting to analyze such constructions we find that the descriptive framework adopted in the book cannot quite handle these. The analysis of examples is also problematic where the *LGSWE* touches upon points in English syntax that, in terms of the given descriptive framework, would allow more than one analysis. Here the grammar usually refrains from taking up one position or another and opts for a form of underspecification. Thus, in the case of participial premodifiers, it is observed that ‘Participial premodifiers, such as those illustrated above, are typically adjectival ...; but in some cases they have the character of noun rather than adjective modifiers, and in yet other cases their word-class membership is unclear ... In this chapter we do not attempt to sub-classify these *-ing* and *-ed* forms as adjectival or nominal, but treat them as a separate category of premodifiers’ (p 575).

One final point I want to make is that I would have appreciated, had the authors included suggestions for further reading at the end of sections or chapters. There are many corpus-based studies that have explored particular linguistic structures or phenomena in great detail, be it on a far more modest scale. It would have been stimulating to have had pointers to such studies immediately in the text, instead of hidden away in the bibliography.

Conclusion

The corpus-based approach adopted in the *LGSWE* brings to light many interesting aspects of language use. One of the more attractive features of this grammar is the attention given to the quantitative dimension of language description. The corpus findings that are presented in this grammar often corroborate earlier accounts that in general were based on far fewer data. Their interpretation in terms of register variation and – less frequently – in terms of dialect differences between British and American English adds to our insights in language structure and the observed variability. Occasionally the corpus findings are tentatively interpreted as pointing to the ongoing process of language change. Typical examples are *no* vs *not* negation and the use of *who* vs *whom*. A point of weakness of the *LGSWE*, I think, is the extent to which this grammar capitalizes on the grammatical description in *CGEL*. The decision to hold on to the descriptive model used in *CGEL* implies that the quantitative description must remain a separate level of description, rather than an integral part of the overall description. Thus, the task to develop corpus-based grammars in which an integral description is given of both the qualitative and the quantitative dimension of language use is left to future projects.

The *LGSWE* comes recommended by scholars such as Randolph Quirk and David Crystal, who predict that ‘the *LGSWE* will rapidly establish itself indispensable’ (Quirk: p v) and ‘[F]or the foreseeable future, anyone with a serious interest in English grammar will have to take into account the information this book contains’ (Crystal: cover blurb).

Note

1. In this case the examples are an illustration of the fact that relative clause constructions in conversation are often quite complex. As the *LGSWE* observes: ‘The existence of such constructions in conversation is surprising, since structural complexity is stereotypically associated with written exposition rather than speech. However, among the types of structural complexity that are characteristic of conversation is the heavy use of complement clauses ... Relative clauses with embedded gaps are a related type of complexity. In formal writing, they would be regarded as awkward at best, whereas in conversation, they are perfectly acceptable and not at all unusual’ (p 623).

References

- Aarts, Jan. 1999. The description of language use. In Hilde Hasselgård and Signe Oksefjell (eds) *Out of corpora. Studies in honour of Stig Johansson*, 3–20. Amsterdam and Atlanta: Rodopi.
- Leech, Geoffrey. 1992. Corpora and theories of linguistic performance. In Jan Svartvik (ed) *Directions in corpus linguistics*, 105–125. Berlin: Mouton de Gruyter.
- Quirk, Randolph. 1968. *Essays on the English language: Medieval and Modern*. London: Longman.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Harlow: Longman Group Limited.