

Book Reviews

Longman Grammar of Spoken and Written English

Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan
(Northern Arizona University, University of Oslo, University of Lancaster, Iowa State University, and University of Southern California)

Harlow, Essex: Pearson Education Ltd,
1999, xxviii+1204 pp; hardbound, ISBN
0-582-23725-4, £72.45

Reviewed by
Graeme Hirst
University of Toronto

1. Introduction

Since its publication in 1985, the outstanding 1,800-page *Comprehensive Grammar of the English Language*, by Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik, has been the definitive description of the grammar of English and an indispensable reference for any research in the analysis or generation of English that attempts serious coverage of the syntactic phenomena of the language. The new *Longman Grammar of Spoken and Written English*, by Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan, is an important complement to the earlier work, extending and sometimes revising the descriptions of Quirk et al., by means of an extensive corpus analysis by the five authors and their research assistants. Now, the bookshelf of any researcher in English linguistics is incomplete without both volumes.

Like Quirk et al. (hereafter *CGEL*), Biber and his colleagues attempt a detailed description of all the syntactic phenomena of English. But Biber et al. (hereafter *LGSWE*) go beyond *CGEL* in several important ways:

- The work is based on corpus analysis, and there is a strong emphasis on linguistic function in the interpretation of the quantitative results of the analysis.
- A central organizing element is the importance of register as a factor in linguistic choices.
- Spoken, conversational English is treated as equal in standing to written English.
- The emphasis is much more explicit than in *CGEL* on the relationships between discourse factors and language users' syntactic choices and between those syntactic choices and language users' lexical choices.

The corpus upon which *LGSWE* is based is the *Longman Spoken and Written English Corpus* of 40 million words of British and American English in 37,000 texts. The corpus includes four main registers: transcribed conversations (6.4 million words),

fiction (5.0 million words), news (10.7 million words), and academic prose (5.3 million words); in addition, there are two supplementary registers: 5.7 million words of nonconversational speech and 6.9 million words of general prose.

The transcribed conversations are of particular note. British and American informants surreptitiously tape-recorded all their conversations for a week. (Other participants in the conversations were asked post hoc for consent to the use of the conversation in the corpus.) Throughout the book, the differences between conversation and the written registers are treated in detail, and Chapter 14 is devoted to particular aspects of conversation, such as how the constraints of real-time, interactive language production influence the language user's choice of syntactic constructions.

While the more mechanical parts of the corpus analysis were largely automatic (beginning with part-of-speech tagging), the main analyses were carried out manually by the authors with the aid of software for searching and manipulating the contents of the corpus. But manual analysis was also required for many of the relatively low-level tasks, where semantic judgment was required to determine the function of a linguistic element. For example, it still requires a human to reliably distinguish instances of the hedging use of *sort of* from instances of the regular noun-phrase use (*you can sort of wedge it in; what sort of ideas have you come up with?* (p. 36)) and to tease out the semantics of the genitive in a coordinate construction (*we present Berg and Muscat's definition* refers to just one definition; but *Andrew and Horatia's eyes met* refers to two separate pairs of eyes (p. 298)).

Table 1 lists the titles of the chapters, along with my gloss of the contents of each. A complete listing of even the first-level subheadings would take much more space.

2. The Analysis

A typical section of *LGSWE* takes a syntactic phenomenon of English, describing it in general terms. It is then subcategorized, each subcategory is explained, and a cross-register corpus analysis of the phenomenon is presented, highlighting its distribution in each register and the distribution of the subcategories. If the subcategories are semantic, a corpus analysis is presented of the various possible syntactic realizations of each. Tables, sometimes quite long, are often given of relevant words and the frequency of some specific behavior. The results are then discussed qualitatively, emphasizing the functional aspects of syntactic choice. Constructions are analyzed with respect to the work or tasks that they perform, their relation to cognitive constraints on language production, and their social indexing (pp. 41–44). The tasks of linguistic features are divided into the ideational, textual (serving to mark information structure or textual cohesion), personal, interpersonal, contextual, and aesthetic. The following quotations exemplify the style of the analysis:

That-clauses functioning as noun complements are one of the primary devices used to mark stance in academic prose. (p. 648)

There is a clear stylistic difference between interrogative *if*-clauses, which are strongly favored in the more colloquial style of conversation and fiction, and interrogative *whether*-clauses, which are more neutral in their stylistic range. (p. 691)

In [academic prose,] fronting serves to juxtapose items which through semantic repetition cohesively tie the sentences together. . . . Conversation and fiction, by contrast, strive for greater impact and stylistic

Table 1
Synopsis of the contents of *Longman Grammar of Spoken and Written English*.

- A. Introductory
 - 1. A corpus-based approach to English grammar
Introduction to the basic concepts of the work, including the use of the corpus.
 - B. Basic grammar
 - 2. Word and phrase grammar
The characteristics of words; the basic word classes; function words; the characteristics of phrases; types of phrase; embedding and coordination of phrases.
 - 3. Clause grammar
The major elements of clauses and their patterns; peripheral elements; ellipsis; negation; subject–verb concord; dependent and independent clauses.
 - C. Key word classes and their phrases
 - 4. Nouns, pronouns, and the simple noun phrase
Types of nouns; determiners; number; case; gender; derived nouns; pronouns.
 - 5. Verbs
Single-word and multiword verbs; semantic domains; valency; auxiliaries; copulas.
 - 6. Variation in the verb phrase
Tense; aspect; voice; modality.
 - 7. Adjectives and adverbials
Types of adjectives; comparatives and superlatives; formation of adjectives; syntactic roles of adverbs; semantic categories of adverbs.
 - D. More complex structures
 - 8. Complex noun phrases
Premodification; nominal sequences; restrictive and nonrestrictive postmodifiers; postmodification by relative clauses, prepositional phrases, and appositives.
 - 9. The form and function of complement clauses
that, *wh-*, *-ing*, and infinitive clauses.
 - 10. Adverbials
Circumstance, stance, and linking adverbials.
 - E. Grammar in a wider perspective
 - 11. Word order and related syntactic choices
Marked word orders, such as inversions; passive constructions; existential *there*; clefts.
 - 12. The grammatical marking of stance
Kinds of stance; attribution of stance.
 - 13. Lexical expressions in speech and writing
Collocations (“lexical bundles”); idioms; free verb–particle combinations; binomial phrases.
 - 14. The grammar of conversation
Differences between conversation and writing; dysfluencies; grammatical characteristics of sentences constructed in real time.
- End matter
- Appendix: Contraction
 - End notes
 - Bibliography
 - Lexical and conceptual indexes

effect, so we find types of fronting which chiefly convey special emphasis and contrast—especially fronting of objects but also some fronting of predicatives. (p. 910).

At the lexical level, there is a strong emphasis on collocation, which the authors generalize to the notion of **lexical bundles**, “sequences of word forms that commonly go together in natural discourse” (p. 990)—for example, *are shown in table, should be noted that, you want a cup of tea, thank you very much, got nothing to do with*. Chapter 13 includes long tables of four- to six-word lexical bundles common in conversation and academic prose, classified by structure.

The authors adopt a prototype view of word classes in the style of Lakoff (1987, Chap. 3) (though it is not described in those terms), with classes having core and peripheral members: “nouns can be more or less ‘nouny’” (p. 59). Nonetheless, the authors categorize words as much as they possibly can, introducing novel classes such as **semi-determiners**: “determiner-like words which are often described as adjectives . . . [but] have no descriptive meaning” (p. 280)—for example, *same, other, former, latter*. This varies from CGEL’s analysis, in which the adjectival aspect of such words is emphasized (e.g., pp. 430–431).

Despite the emphasis on frequency data, there are few actual numbers. In the text itself, most of the results are given qualitatively, with just occasional, isolated numeric data points that are usually approximations. For example, the past perfect is compared with the simple past tense in these three bullet points (p. 469):

- The past perfect aspect has accompanying time adverbials a greater percentage of the time than does the simple past tense.
- Past perfect verb phrases often occur in dependent clauses.
- Taken together, these two factors (time adverbials and dependent clauses) account for c. 70% of all occurrences of past perfect verb phrases.

(Statistical significance is never noted; the authors promise in the introduction (p. 40) that anything reported is statistically significant.) Even the tables and figures are rarely numeric. Most quantitative data are reported as histograms, but the histograms, either by their small size or by their design (see below), are very low in precision and are apparently intended only for quick visual impressions.

A representative section is 10.2 (pp. 776ff.), on circumstance adverbials. The first subsection describes the different semantic categories of these adverbials (place, time, means, etc). The second subsection gives corpus results on the frequencies of each of the categories in each of the four registers of the corpus, with discussion of the functional significance of the findings; for example, “News is particularly concerned with current events. Time adverbials are therefore commonly used to make clear when events happened or to give background leading up to the current event” (p. 785). The next three subsections give the frequencies of different kinds of phrasal and lexical realizations of each of the semantic categories, including a 2.5-page table of the frequencies in each of the registers of the most common lexical realizations. Differences between British and American usage are highlighted; for example, a large difference is found in the choice between relative and absolute references to time in the news register: *yesterday* is eight times more common, and *this morning* is more than twice as common, in the British news subcorpus than in the American news subcorpus, whereas names of days of the week are 28 times more frequent in the latter. The next two subsections look at the frequencies with which the circumstance adverbials occur

in different positions within the sentence and the correlation between their position, semantic category, and length, and then at the order of occurrence of multiple circumstance adverbials within the same sentence. The final subsection, which is almost as long as all the others together, repeats much of this analysis for clausal realizations of circumstance, which the authors find to be different enough from phrasal and lexical realization to warrant a separate presentation of the analysis. In total, this section is 77 pages long.

The authors are generous with the use of examples from the corpus to illustrate their descriptions. Often, the selection of examples is as large as, or even larger than, the description itself; for example, Section 10.2.1.2, on circumstance adverbials of time, consists of five sentences by the authors and 14 example sentences from the corpus. And as anyone who has worked with corpora knows, it is sometimes hard to avoid the distraction of reading the data for its own sake; similarly, in *LGSWE*, it's easy to start skipping over the authors' text in order to hurry to the next example. One suspects the authors of having sometimes deliberately chosen the most enticing examples from the corpus to illustrate each point; I got particularly hung up trying to construct a story around one example on page 809:

One day I left the Yale on the latch accidentally, and when I came back
I found a brand-new shelf, with a brass rod below it, high up in the
shallow recess beside the fire.

Other examples highlight the drama of daily life (pp. 1104, 746, 1119):

A: Can I have a drink?
B: Yeah, what would you like?
A: I would like strawberry.
B: With or without ice?
A: With.

She keeps smelling the washing powder.

Yeah, he went "Oh!" He goes, "Who put that there?" And the bit where he goes he goes "Urgh, cobwebs", and she goes "Piss off!" She goes "Mum, come and sit here", she goes "Piss off!" like that and the mother goes "You talking to me?"

3. Empiricism versus prescriptivism

Since many lay readers of *LGSWE* will turn to the book for guidance in "correct English," the authors are careful to emphasize that their work is descriptive, not prescriptive. Of course, this is not to say that the book is not of use as a guide for those who are uncertain in their usage; any writer who wants to ensure that his or her usage is in accord with English norms, be they prescriptive or statistical, will find it extremely helpful. For example, the section on verb concord with existential *there* (p. 186) spends some time discussing the fact that the prescriptively deprecated use of a singular verb followed by a plural noun phrase (*There's apples if you want one*) is actually more common in conversation than the prescriptively approved form (*There are apples . . .*). But there is less emphasis than in *CGEL* on the identification of prescriptively deprecated or stigmatized forms. Indeed, whenever prescriptivism is mentioned in *LGSWE*, a certain tension arises. For example, the section on dangling participles

(p. 829) first mentions the prescriptivist proscription of such participles as if preparing to discard it, but then concedes that violation of the rule can lead to absurd interpretations; and the examples offered are constructed rather than selected from the corpus, implying that none were found in the corpus and hence that dangling participles have no part in even a purely descriptive grammar. (The section closes with some corpus examples of unattached participles that don't "dangle" in the proscribed manner.) In another round of prescriptivism-bashing on pages 83–84, we read:

There is a well-known prescriptive reaction against beginning an orthographic sentence with a coordinator [such as *but*]. Nevertheless, in actual texts, we quite frequently find coordinators in this position. . . . The prescription against initial coordinators seems most influential in academic prose. The higher frequencies in fiction and news reportage probably reflect the fact that these registers often include more spontaneous discourse, including fiction dialogue and quoted speech in news, evidencing the lack of attention to prescriptive rules of ordinary speech.

But while plausible, this is just speculation, and is not in keeping with the authors' own stated goal of explaining their corpus observations in terms of linguistic function; it simply cannot be determined from the corpus data whether or not prescriptivism influences the observed tendency of academic writers to avoid sentence-initial coordinators. (It certainly didn't influence my writing of the previous sentence.¹)

4. Pragmatics

LGSWE far exceeds *CGEL* in the quality of its index, which was prepared by Meg Davies. The authors have learned from the problems of *CGEL*'s enormous index (by David Crystal), which, while very comprehensive, frustrates the user by being very sparing with its subheadings. It is not unusual for a single *CGEL* index entry or subentry to list more than 50 locators without differentiation; moreover, the locators are paragraph numbers, not page numbers, although the design of the book makes it hard to find the paragraph numbers when flipping pages. By contrast, the index of *LGSWE* uses descriptive subheadings for every small group of locators—and the locators are page numbers. In addition, the lexical index is separate from the conceptual index, which also adds to ease of use.

LGSWE is an attractive, well-designed, and well-typeset book. I noticed no significant typos or errors of copyediting. But I must admit to being both mystified and somewhat annoyed by the histograms. Five different styles are used, all of them low in resolution: cumulative vertical bars, adjacent vertical bars, horizontal bars of a few discrete lengths, horizontal rows of squares, and horizontal rows of circles. The vertical styles require the reader to distinguish up to six shades of grey. The horizontal styles have no scale, but rather a legend such as "each ■ represents 5%," so that one has to count the squares or circles. Confusingly, the style chosen for any particular histogram seems often to be random and unrelated to either the nature of the data shown or the logic of the display. For example, Tables 5.26–28 (pp. 432–3) show the distribution across registers of various forms of *do*; the first two tables use horizontal

¹ When I was young, my grandmother admonished me that the word *but* must always be preceded by a semicolon; but it was clear that she herself did not agree with this rule, and was passing it on to a new generation merely as her grandmotherly duty.

squares, but the third uses horizontal circles. And then Table 5.31 (p. 437) shows similar data—frequencies of common copular verbs across registers—by means of terms such as “rare” and “relatively common”: a verbal histogram! A similarly disparate group appears on page 784: Figure 10.4 shows the raw frequencies of semantic categories of circumstance adverbials across the registers as cumulative vertical bars, whereas Figure 10.5, immediately adjacent, shows the percentage data for one subcategory of these adverbials as adjacent vertical bars, and Table 10.3, just below, shows similar data for a different subcategory as horizontal rows of squares. (Moreover, Table 10.3 is captioned in quite different terms from Figure 10.5, although the two are in fact logically parallel.) I spent a while puzzling over what the difference was supposed to be in meaning, implication, or function between all these styles, but finally had to conclude that there was none.

The book seems to be well bound, and comes in a slipcase to help protect it. Again, this contrasts with *CGEL*, whose binding was inadequate for heavy use of a heavy book; my copy fell part after five or six years of repeated use and loans to students with backpacks, and had to be re-bound.

The publicity brochure that came with the review copy of *LGSWE* gave its price as £69.00, but as this review is written, the price given on the Longman Web site is £72.45; Amazon lists it for \$119 (and *CGEL* for \$247). This is not expensive for an academic book, especially one of this size and one in which the publisher has invested so much in resources, but it is sufficient to keep it out of reach of many students and others who would benefit from owning their own copy. By contrast, the hardback edition of the 1670-page *Longman Dictionary of Contemporary English* (third edition with CD-ROM) is only £20.88 (\$39 at Amazon).

5. Conclusion

Despite its size and all the work that went into it, *LGSWE* is not a replacement for *CGEL* but rather a complement to it; indeed, the authors explicitly describe it as such (p. viii). In particular, the descriptions of English grammar in *CGEL* are perhaps twice as extensive as those of *LGSWE*. (*CGEL* is about 1.5 times larger than *LGSWE*, and yet has far fewer examples and no corpus analyses.) This is not to say that *LGSWE* doesn't stand on its own. It does, and for many people, it is the only book of English grammar that they need. In fact, many students will get by just fine with one of the many less-detailed and more-concise books available, such as those of Wardhaugh (1995), Baker (1995), Greenbaum (1996), or McCawley (1998) or the student abridgement of *CGEL* (Greenbaum and Quirk 1990). But the most serious users, including many readers of this journal, will now wish to consult both *LGSWE* and *CGEL* on any question of the more subtle aspects of English grammar and lexis.

Acknowledgments

I am grateful to Suzanne Stevenson and Nadia Talent for helpful comments on earlier drafts of this review. Preparation of the review was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

References

Baker, C.L. 1995. *English Syntax* (second edition). The MIT Press, Cambridge, MA.

Greenbaum, Sidney. 1996. *The Oxford English Grammar*. Oxford University Press.

Greenbaum, Sidney and Randolph Quirk. 1990. *A Student's Grammar of the English Language*. Longman.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. The University of Chicago Press.

McCawley, James. 1998. *The Syntactic Phenomena of English* (second edition). The University of Chicago Press.

Quirk, Randolph, Sidney Greenbaum,
Geoffrey Leech, Jan Svartvik. 1985. *A
Comprehensive Grammar of the English
Language*. Longman.

Wardhaugh, Ronald. 1995. *Understanding
English Grammar: A linguistic approach*.
Blackwell, Oxford.

Graeme Hirst is the book review editor of *Computational Linguistics*. His research topics include lexical and syntactic indicators of style, the problem of near-synonymy in machine translation, and computational models of linguistic context. Hirst's address is Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 3G4; e-mail: gh@cs.toronto.edu; URL: <http://www.cs.toronto.edu/~gh>.