

Hans Lindquist and **Magnus Levin**. *Corpus linguistics and the description of English* (Edinburgh Textbooks on the English Language – Advanced). Edinburgh: Edinburgh University Press. 2018. Second edition. 256 pp. ISBN 9781474421720. Reviewed by **Stefan Diemer**, Trier University of Applied Sciences.

Hans Lindquist's *Corpus linguistics and the description of English* has been a valued companion textbook for students of English since its publication in 2009. In its new second edition and co-authored with Magnus Levin, this contribution to the 'Advanced Edinburgh Textbooks on the English Language' series succeeds in maintaining its leading position among the existing introductions to corpus linguistics. It also is an excellent read for both linguists and non-linguists. While keeping the structure of the original edition, the volume has been extensively reworked to include new trends particularly in the area of web-based corpus linguistics. It also features recent research, updated corpus references and applied examples, making it an eminently suitable textbook not only for linguistics or literature majors but also for students of other disciplines.

Lindquist and Levin start the book with a disclaimer – they present corpus linguistics not as one of the “hyphenated branches” (p. 1) of linguistics, but as an efficient methodology to study language in use. In their introduction they emphasize “the joy and fascination that lie in the description and analysis” (p. 1) of language, and this passionate view is clearly evident in the ten chapters that make up this volume. The authors cover a wide range of topics, from the origins and motivation of the corpus approach to recent trends in web-based corpus research and corpus creation, liberally annotated with corpus information, study questions, exercises, and DIY sections. Two generations of corpus linguists teamed up for this edition, and the authors' profound expertise in research and teaching is clearly in evidence throughout, both in the concise discussion of key issues and the thoughtful methodology of exercises. Through its systematic structure, the book can easily and profitably be used as a week-by-week semester textbook or, alternatively, provides self-contained study units in specialist seminars that seek to integrate a corpus perspective.

The two preparatory chapters 1 and 2 cover basic corpus terminology and methodology and thus provide an excellent introduction in itself. In Chapter 1, Lindquist and Levin start with an overview of corpus linguistics' history and a description of the various types of corpora, extending the discussion to very large and small corpora in view of latest developments. They ask, "what can we get out of corpora?" (p. 5) and proceed to demonstrate the key advantages of speed and reliability in describing language that only corpora can deliver. The authors also place corpus linguistics solidly in today's linguistic mainstream. For Lindquist and Levin, there is no conflict between other linguistic disciplines and corpus linguistics, as corpus methods in the form of data sets and research tools are widely accepted across all linguistic disciplines. Their discussion of Svartvik's (1992) famous arguments helps readers understand that corpus linguistics provides a wide range of methods in the spectrum between corpus-based and corpus-driven approaches, and that all corpus research is by no means quantitative, illustrated by the continuing need for careful qualitative analysis even as big data becomes more pervasive. A description of corpus types, with special emphasis on the complications of spoken corpus compilation, concludes the chapter. Chapter 2 focuses on a description of the two main approaches, qualitative and quantitative. The authors also discuss issues of representativity and annotation. Lindquist and Levin favor a combination of quantitative and qualitative methods, reflecting the increasingly widespread view in the corpus linguistics research community (and, for that matter, Chomsky's original pet peeve with the emerging method) that frequency data is of limited use on its own, and that it is vital, for a good corpus study, "to look at the reality behind the figures" (p. 26). This approach ties in with their nuanced view of statistical significance. While the authors echo Stefan Gries' (2006, 2015) repeated call for greater statistical sophistication, they also stress that experts do not agree which statistical measures are most suitable for corpus studies. They mention chi square (and its issues with very small and very large datasets) and lexical association testing, though one would have wished for a brief discussion of the increasingly popular regression models. The issue of representativity and its limiting effect on our ability to generalize corpus results is also emphasized. Chapter 2 concludes with a short overview of tagging and annotation, though a detailed discussion of pragmatic, discourse and prosodic annotation is conspicuously missing.

Chapters 3 to 9 systematically illustrate specific areas of English from a corpus perspective: words, collocations, phrases, metaphors, syntax, gender, and language change. Through their combination of theory, discussion, examples and exercises, these are ideally suited for a semester program, particularly as the

authors can draw upon their own extensive experience and research associations to contextualize their considerations, creating a rich portfolio of applied studies. Chapter 3 focuses on the key application (and initial main selling point) for corpus linguistics, enabling lexicographers to collect “large numbers of authentic examples in context” (p. 69) in order to make informed judgments about the compilation of dictionaries. The authors discuss Sinclair’s (1998) ‘semantic prosody’ to illustrate “hidden evaluative meanings” (p. 70) and provide illustrations of using corpus analysis of lexis to document language change and identify register and genre differences. Chapters 4 and 5 deal with combinations of words, divided into collocations and phrases. The authors present a detailed collocational study illustrating the use of corpus methodology as a tool for enhancing language learners’ production and teachers’ lesson preparation and grading. They also discuss mutual information, z-score and log-likelihood as statistical measures for collocational strength, cautioning that absolute frequencies may yield a very different picture. Again, the importance of qualitative analysis is emphasized. The study of collocations is supplemented by the analysis of longer phrases or formulaic sequences in Chapter 5. Lindquist and Levin point out that contrary to traditional research in phraseology, corpus studies have shown that ‘fixed’ expressions are variable and proceed to give examples from corpora such as the British National Corpus (BNC) and the London-Lund-Corpus, as well as using the Web as a corpus. A second focus is the study of recurrent strings (n-grams), whose broad potential is exemplified by an n-gram study of the BNC and a stylistic study of clusters in novels (Mahlberg 2007). The authors stress the important role of phrases for language processing and production as well as reception and conclude that “language is a mixture of repetition and creative innovation” (p. 108). The authors next turn to a major issue in corpus linguistics – how to study meaning, in particular metaphorical or metonymical meaning with the help of a corpus. Though Chapter 6 seems slightly isolated between the discussion of phrases and the following focus on grammar, the applications of corpus methods for the study of meaning are of considerable interest in a study setting. The key issue remains the identification of metaphorical and metonymical meaning, for which the authors propose a lexical approach, starting with source or target domains in combination with a collocation analysis. A brief discussion of automated semantic tagging, for example with the CLAWS semantic tagger, would have been useful here, but the untagged examples and exercises that the authors present provide a good basis for getting students started in the field. In Chapter 7, the authors give examples for grammatical features that can be studied in and through corpora. They focus on morphological features and grammar but expand their approach to the lexicogrammatical interface, e.g. in

adjective or noun complementation, with an increasingly abstract search string design and also using parsed corpora. Here, Lindquist and Levin admirably succeed in emphasizing the need for a flexible, trial-and-error approach in order to optimize search strings and produce reliable results, as well as the variety of available corpora. For Chapters 8 and 9, the authors select a sociolinguistic feature, gender, and a diachronic feature, language change, to be illustrated in terms of a corpus approach. In Chapter 8, they investigate markedness in gendered nouns, but also how men and women are described, and what actions they perform in a corpus context such as the BNC or the diachronic Corpus of Historical American English (COHA), in order to illustrate how social features are reflected in language. They also comment on some features that have been described as marked in gendered discourse, such as the use of color terms and hedging. While their results mirror some past research (as in color use), the authors stress the importance of considering the corpus context and other potentially interfering social factors, as well as the issues of frequency and representativeness, concluding that “with complex discourse issues qualitative analyses are called for” (p. 172). Essentially, they caution that while you may describe men’s and women’s language use and ongoing changes through corpora, the relation between these findings and the societal reality is a matter of opinion. In Chapter 9, Lindquist and Levin turn to language change in a corpus context. They illustrate the use of corpora to investigate and document language change with a historical example, the grammaticalization of *beside(s)*, and an example illustrating current language change with adverbial *likely*. In a diachronic context, corpus composition must be carefully considered, not only in terms of size, but also in terms of text type, genre and genre development. The authors also present diachronic and historical corpora that are available to investigate language change, but also how synchronic corpora can be used for diachronic research, e.g. by documenting a particular situation and then comparing it to other corpora.

The last chapter, 10, focuses on new developments and illustrates how corpus linguistics is used to investigate language in cyberspace. As counterpart to the two-chapter introduction, this chapter provides an ideal conclusion to this introduction to linguistics, pointing towards a future where more and more corpus research is done on the web. The authors discuss how the web can be used both as a corpus and for corpus compilation and caution that the web as a corpus approach can be a useful complement to searches in other corpora, but that the data itself is problematic due to its informal, unedited and unstructured nature. They illustrate the potential of this approach with a few basic examples, also mentioning the possibility to study sociolinguistic features (such as dialects) on

the web. A discussion of the web as a useful source for corpus compilation concludes the volume.

A textbook necessarily needs to be selective in its representation of the respective field in order to provide a general view, but also to depict the range of potential applications. While the volume covers a wide range of topics, it would have been nice to read more about the use of corpora for pragmatics and discourse analysis, automated tagging or parsing, or to discuss new ways of compiling spoken corpora through automated transcription. In view of the increasing importance of social media and computer-mediated communication in general, an overview of corpus studies in this field might have been added, replacing some of the original studies in the volume as part of the revision for a second edition. As developments tend to rapidly overtake publications in these fields, though, perhaps this will be better left to a separate volume on corpus linguistics on the web. As the authors formulate in their introduction: “The ultimate aim is to learn more about the language and to understand better how it works.” (p. 1) In this aim, Lindquist and Levin succeed admirably, and at the same time they have created an accessible, varied and proficient guide to corpus analysis that is both immensely readable and ideally suited as a students’ textbook.

References

- Gries, Stefan T. 2006. Exploring variability within and between corpora: Some methodological considerations. *Corpora* 1 (2): 109–151.
- Gries, Stefan T. 2015. Quantitative designs and statistical techniques. In D. Biber and R. Reppen (eds.), *The Cambridge handbook of English corpus linguistics*, 50–71. Cambridge: Cambridge University Press.
- Mahlberg, Michaela. 2007. Clusters, key clusters and local textual functions in Dickens. *Corpora* 2 (1): 1–31.
- Sinclair, John. 1998. The lexical item. In E. Weigand (ed.), *Contrastive lexical semantics*, 1–24. Amsterdam: John Benjamins.
- Svartvik, Jan (ed.). 1992. Corpus linguistics comes of age. In J. Svartvik (ed.), *Directions in corpus linguistics. Proceedings of Nobel Symposium 82 Stockholm, 4–8 August 1991*, 7–14. Berlin: Mouton de Gruyter.