Review of *Corpus linguistics for vocabulary: A guide for research*

Rebeca Arndt, University of Central Florida

---

**Corpus linguistics for vocabulary: A guide for research**

Szudarski, P.

2017

ISBN: 9781138187221

US $ 41.95

228 pp.

Routledge

New York, NY

---

**Introduction**

Szudarski’s *Corpus linguistics for vocabulary: A guide for research* is an excellent guide for undergraduate and graduate students, teachers, and researchers, whether they are novices or experts in the fields of corpus linguistics and corpus analysis. The author delivers a practical manual on a topic some perceive as overly technical by presenting the content in a very comprehensible manner. From the very first pages, this guide enables educators unfamiliar with the concept of a corpus to feel confident about exploring its various technological tools and methods.

In the introduction, the author states his intention to guide the readers to research *corpora*, defined as “principled and computer-readable collections of texts that allow analysis of patterns of language use across different contexts” (p. 3). This definition—informed by that of Biber, Conrad, and Reppen (1998)—includes the concept *principled*, yet avoids terms such as *qualitative* and *quantitative*, illustrating the author’s preference for using clear language that does not discourage practitioners. The volume is organized as a collection of 10 chapters, each of which includes a convenient summary and reference list that can be employed to further examine the topic of the chapter.

Chapter 1 introduces the topic of corpus linguistics and first establishes the differences between corpus, corpus-based, and corpus-driven analyses. The chapter then defines different corpora categories (e.g., general and specialized), while referencing supplementary tools, examples, and tutorials, many of which are Internet-based and free of charge. These instruments include the British National Corpus, the Contemporary Corpus of American English, the Michigan Corpus of Academic Spoken English, and the English as a Lingua Franca in Academic Settings Corpus.

Dedicated to corpus analytic tools and statistical tests, Chapter 2 provides the reader with the means to conduct small-scale corpus analyses. In terms of tools, a clear distinction is made between web-based and computer-based applications (concordancers). After introducing several tools, the chapter presents an exploratory framework for conducting research. The first tutorial comprises a frequency analysis to gain insight into the nature of authentic communication, introducing technical terms such as *chunks*, *n-grams,*
lexical bundles, and keywords.

Chapter 3 turns the reader’s attention toward vocabulary. In the first section of the chapter, the author offers clear definitions of words, vocabulary, and lexis, as well as of three related key terms in corpus linguistics: lexeme, lemma, and word family. To introduce the complex construct of vocabulary knowledge, Nation’s (2001) three-part categorization of word knowledge is used: form, meaning, and use (both receptive and productive). One especially good feature of this chapter is the inclusion of examples of research questions and exemplary projects focused on vocabulary.

Particularly helpful for practitioners, Chapter 4 focuses on the importance of frequency differences between function words and content words by discussing the concepts of lexical coverage and vocabulary bands. The chapter introduces the reader to various vocabulary lists (e.g., the New General Service List, the Common Core List, and the Academic Word List), as well as to tools like VocabProfile and AntWord Profiler (lexical frequency profiling platforms used to analyze vocabulary frequency and word distribution range.

Chapter 5 delves into phraseology as a fundamental quality of productive language. The chapter begins by presenting the theory of lexical priming featuring co-occurrences (i.e., collocations and colligations) as a major niche in corpus research. It then moves on to cover formulaic language, lexical bundles, clusters, and multiword units. One of the take-aways from this chapter is the PHRASE List, a phraseology index containing “non-transparent phrasal expressions” (p. 91), which may prove handy to both instructors and learners of English as a second language.

Chapter 6 covers the impact of corpus linguistics on teaching vocabulary. While corpora indirectly serve language pedagogy as a source of teaching and assessment materials, they can also be used directly through a data-driven learning approach. This approach directly involves learners in the process of corpus analysis. Chapter 6 also includes a remarkable selection of web-based corpus resources for exploring vocabulary.

Chapter 7 discusses learner language and characteristics of learner corpora, focusing on how they can inform language assessment and materials development. Of great interest for practitioners is discussion of contrastive interlanguage analysis, a corpus analytic approach for comparing specific features of language produced by learners from different populations. Equally informative is the discussion of corpora containing English as a lingua franca (e.g., the Vienna–Oxford International Corpus of English). The chapter also discusses the English Profile Project, a useful resource helping teachers identify which of six sublevels of proficiency a word, phrase, or grammatical item matches on the Common European Framework of Reference for Languages.

Chapter 8 focuses on specialized corpora connected to vocabulary (e.g., language for specific purposes; LSP), with special attention given to register and genre analysis. Because most LSP corpora are in English, the chapter also specifically addresses English for specific purposes. Arguably, the most valuable components here are introductions to and discussions about the Academic Word List, the Academic Vocabulary List, the Academic Formulas List, and the Academic Collocation List.

In Chapter 9, the focus shifts to discourse, discourse analysis, and lexical features of discourse pragmatics (language in context). First, corpus-based analyses of vocabulary are examined through the lens of discourse, addressing the phenomena of relexicalization and intertextuality. Second, the author discusses corpus pragmatics, defined as a blend of corpus linguistics and pragmatics. The chapter concludes with an interesting illustration of how pragmatic meaning can be gleaned from semantic prosody analysis, with an analysis of the Bank of English corpus showing how the phrase true feelings may be associated with conceal or hide.

The final chapter of the book ends with a summary of the volume and a detailed discussion about future research in corpus linguistics. Szudarski suggests incorporating corpus tools and methodologies with teaching and research, using multimodal corpus approaches, and conducting methodological triangulation. Additionally, he offers several ideas for corpus-based projects that readers can undertake (e.g., analyzing
the use of phrasal versus single-word verbs in American English), each including a research question, background information, and suggested methodology. Graduate students looking for ideas for their own research may find this chapter especially helpful.

In this volume, Szudarski examines essential topics pertaining to corpus-based vocabulary studies such as word frequency, idioms, lexical chunks, phraseology, and pragmatics. Although these topics have certainly been written about previously, this book brings forth new information about current research, tools, and practices developed, particularly in the last decade, and aims it directly at teachers and students. They will find the information about many cost-effective (i.e., mostly free) and user-friendly corpora and corpus analysis tools extremely useful.

Although I believe Szudarski’s book makes a brilliant contribution to the field of corpus linguistics by being up-to-date and accessible, the word research in its title may lead some practitioners to overlook it. However, I believe that classroom teachers would like the book if they see beyond its title. They would learn to appreciate what the field of corpus linguistics can do for their practices. Teachers equipped with the knowledge and tools presented in this book would be ready to conduct small scale, classroom-relevant experiments. The book provides a very accessible source of knowledge for corpus linguistics by seamlessly organizing discussions of theory, research, and technology. Teachers who read this excellent work may be inspired to pursue their own corpus research, guiding them as they enhance their vocabulary instruction with corpus-informed insights.

**References**


**About the Author**

Rebeca Arndt is pursuing a PhD in Education (TESOL track) while working as a Graduate Teaching Assistant and Associate at the University of Central Florida. She instructs undergraduate courses (e.g., *Theory and practice of teaching ESOL students in school* and *Academic language and literacy development*).

**E-mail:** rebeca.arndt@ucf.edu