Corpus Linguistics for Vocabulary: A Guide for Research, 1st Edition. Pawel Szudarski. New York, Routledge (2018). X + 238 pp., ISBN: 978-1-138-18722-1 (pbk).

Muhammad Hossein Parvaz [a, *]

[a] Urmia University, Iran

Corpus linguistics is the study and analysis of data obtained from a corpus. The field has been developing considerably in the last two decades due to the great possibilities offered by the processing of natural language with computers. Different types of books on corpus linguistics are available on the market. However, the need for vocabulary corpus studies motivated Szudarski to develop *Corpus Linguistics for Vocabulary*. It aims to provide a practical introduction to using corpus linguistics in vocabulary studies. Using freely available corpus tools, the author intends to provide a step-by-step guide on how corpora can be used to explore key vocabulary-related research questions and topics such as: what the frequency of each of word in English is and how one can choose which words to teach to learners; how spoken vocabulary differs from written, and how academic vocabulary differs from general; and how vocabulary contributes to the structure of discourse and the pragmatic functions it fulfils.

This introductory volume aims to provide a clear and accessible guide for students, teachers and researchers who have little or no experience in corpus linguistics and want to understand, appreciate and conduct corpus-based research in vocabulary studies. It highlights the usefulness of corpora for studying vocabulary by illustrating case studies and tasks. As emphasized throughout all the chapters, vocabulary is a highly active field with a variety of issues that can be explored by means of corpus tools.

The book consists of an introduction, 10 chapters, a glossary, a commentary on tasks, and a [subject] index. The chapters can be divided into two parts. The first part, Chapters 1-2, provides a brief overview of corpus linguistics as a whole, while the second, Chapters 3-10, introduces vocabulary studies and addresses more specific issues in this area.

* Corresponding author: Urmia University, Iran
Email address: mh.parvaz@urmia.ac.ir

Chapter 1 introduces corpus linguistics as a methodology which relies on corpora. In addition, the chapter discusses the usefulness of corpus-based analysis and its limitations. Finally, it provides an overview of different types of corpora, focusing in particular on databases freely accessible via web-based interfaces. The following chapter, Chapter 2, presents a wide range of corpus tools and statistical tests commonly employed by corpus linguists. The chapter emphasizes on both the ability to use these tools which is a necessary condition to undertake different types of corpus-based research and the importance of mixing quantitative and qualitative types of corpus analysis.

Chapter 3 highlights the importance of vocabulary as a factor influencing general proficiency and constituting a key element of becoming a mature language user. Understanding how vocabulary is learned and used by individual speakers and writers is of paramount importance to the field of applied linguistics as a whole. This chapter introduces the area of vocabulary studies and discusses key terms such as type, token, lemma and word family. It also demonstrates that vocabulary is a multifaceted construct that should be approached from different linguistic perspectives so that its complexity can be fully captured.

Chapter 4 underlines the importance of frequency for corpus research. The chapter explains how frequency-based analyses of word occurrences reveal differences between spoken and written vocabulary. It also discusses the usefulness of frequency when it comes to categorizing vocabulary and creating lexical profiles of individual texts. Finally, it presents how information about the occurrence of words can be fully employed by language teachers and materials developers.

Chapter 5 introduces phraseology and underlines the usefulness of corpora for exploring the nature of formulaic language. Next, the chapter demonstrates how corpus-based analyses can be employed to identify and classify various kinds of lexico-grammatical phraseological units as important elements of naturally occurring discourse. Finally, it shows how corpora are instrumental in studying phraseological patterning and the way meaning is created and conveyed by means of multiword units.

Chapter 6 addresses the impact of corpora on teaching vocabulary and discusses the indirect and direct applications of corpus data. With reference to the former, the chapter presents the importance of corpus data for the development of teaching syllabuses, textbooks, dictionaries, pedagogical wordlists and online resources. As regards the direct use of corpus data in the language classroom, the pros and cons of data-driven learning are discussed. Finally, it demonstrates the potential of purpose-built pedagogic and teaching-oriented corpora.

Chapter 7 discusses the usefulness of learner corpora as a powerful tool for studying learner vocabulary. The chapter starts with describing features of such corpora and types of methodologies that can be applied to analyze learner data. Then, it presents a number of ways in which learner corpora can facilitate research into L2 lexical learning as well as tracking L2 learners' lexical development over time. Finally, it explains how learner corpora can be usefully applied to assess learner language and develop teaching materials.

Chapter 8 underlines the importance of specialized corpora in the study of vocabulary. First, the chapter discusses the characteristic features of such corpora. Next, it explains how they can be employed to study specific uses of vocabulary and the lexical variation they entail. Last, to demonstrate the usefulness of specialized corpora, it presents examples of their applications in a range of areas.

Chapter 9 focuses on the use of corpora for exploring vocabulary at the level of discourse. First, the chapter presents the benefits of combining corpus and discourse approaches for the purpose of analyzing the use of words and phrases. Next, it discusses the importance of discourse-oriented corpus research for studying the role of lexical features in constructing and organizing discourse. Finally, it demonstrates how corpus techniques can assist corpus researchers in analyzing pragmatic functions of vocabulary use.

Chapter 10 serves as a summary of the book and offers a number of ideas for research projects to be undertaken by individual students and teachers, such as frequency-based analyses of texts, differences between spoken and written vocabulary, the role of collocations and other types of phraseological units or register- and genre-specific descriptions of vocabulary use.

I would recommend this simple, informative and practical volume to anyone interested in corpus studies in vocabulary. The writer has tried to fill the gap in the literature and to cover almost all of the concepts and issues related to studies on vocabulary corpus clearly and effectively. A strong feature of the book is the inclusion of many useful figures and tables that capture the research findings in a concrete manner for the reader. Also practical activities and tasks familarize the reader with different aspects of corpus analysis. This excellent book should be a required reading for students and teachers involved in corpus-based research and is generally useful to anyone who seeks a more comprehensive understanding of the resurgence of empirical linguistics.

However, I would like to make some suggestions to improve its readability and to make the volume more interesting and thorough. First, a list of abbreviations can be included as Jones and Daniel (2015) have done for a similar volume for grammar corpus. Second, the Contents can include second-level headings or section titles within the chapters as well for an easy reading and reference. Third, it can raise issues or topics for discussion at the end of each chapter to make the reader critically think about different issues which might not have been discussed thoroughly. Fourth, it needs to mention and explain which type of research method related to corpus linguistics has been used in the volume to trace a path from data to theory. For example, it can refer to Wallis and Nelson's (2001) 3A model: Annotation, Abstraction and Analysis. Annotation consists of the application of a scheme to texts. Abstraction consists of the translation of terms in the scheme to terms in a theoretically motivated model or dataset. Analysis consists of statistically probing, manipulating and generalising from the dataset. The book reviewed here seems to have used analysis.

In general, I can claim this reader-friendly volume has achieved its purpose and made a very practical contribution to corpus linguistics.

References

Jones, C., & Waller, D. (2015). *Corpus linguistics for grammar: A guide for research* (1st ed.). New York, Routledge.

Wallis, S., & Nelson, G. (2001). Knowledge discovery in grammatically analysed corpora. *Data Mining and Knowledge Discovery*, 5: 307–340.