



## Review of *Corpus Linguistics for Vocabulary: A Guide for Research*

Amare Tesfie Birhan <sup>a</sup> \* 

<sup>a</sup> Bahir Dar Institute of Technology, Bahir Dar University, Ethiopia

Received 22 January 2019 | Received in revised form 31 May 2019 | Accepted 28 September 2019

---

### APA Citation:

Birhan, A. T. (2019). Review of corpus linguistics for vocabulary: A guide for research. *Eurasian Journal of Applied Linguistics*, 5(3), 543-545. Doi: [10.32601/ejal.651351](https://doi.org/10.32601/ejal.651351)

---

### Abstract

*Corpus Linguistics for Vocabulary: A Guide for Research*. Pawel Szudarski. London/New York: Routledge (2018). X+228, PP. ISBN 978-1-1-138-18722-1. £29.99.

© 2019 EJAL & the Authors. Published by *Eurasian Journal of Applied Linguistics (EJAL)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** English language teaching; corpus linguistics; vocabulary

---

Corpus linguistics is one of the newly emerging instructional methods in the field of language teaching and learning. Corpus gives the opportunity to use authentic language data rather than contrived or made up examples for classroom pedagogy and language study; learners, researchers, material writers, and syllabus designers can use authentic vocabularies, collocations, and linguistics patterns that are frequently used in a specific communicative contexts by the help of computer software programs (Yoon & Hirvela, 2004). This has been attracting researchers and practitioners since 1950s.

Considering the wider need of language teachers and linguists on corpus linguistics, the author Pawel Szudarski contributed to a book entitled '*Corpus Linguistics for Vocabulary: A guide for research*,' with the aim of providing a practical guide on how to use corpus for vocabulary studies. The author has approached the ideas by focusing on the eight issues that are commonly raised by researchers and language teachers such as content and function words, the difference between spoken and written vocabulary, collocation and multiword units, using corpus data for language study, features in learner vocabulary, academic vocabulary and general vocabulary, vocabulary to structure discourse and pragmatic function of vocabulary. This introductory volume is aimed to be a practical guide for students, teachers and language practitioners with no experience in corpus analysis. The ideas have

---

\* Corresponding author.

E-mail address: [amaretesfie@gmail.com](mailto:amaretesfie@gmail.com)

presented clearly by supporting with numerous practical examples on how to use and study vocabulary. Accordingly, the book is composed of ten chapters.

Chapter one explicates the idea of corpus linguistics as methodology. The author discusses the major issues that are considered during corpus design such as representatives, size and adequate coverage of the sample language data. Szudarski also explains the benefits and limitations of corpus analysis. In this chapter, the author introduces different types of web-based corpora that can be used for corpus-based research. Chapter two focuses on corpus analysis tools and statistics. The author describes different types of corpus analysis tools such as frequency analysis and concordancing, wordlists, word combinations, and n-gram analysis. Similarly, the author introduces the most common statistical tests in corpus linguistics such as t-tests, Log-likelihood, T-score and mutual information. The importance of mixed (quantitative and qualitative) corpus analysis is given due emphasis.

Chapter three defines vocabulary and related terminologies. This chapter elaborates the importance of vocabulary in language teaching and learning. It is explained that vocabulary is the core language skill. The author claims that there is a variation in vocabulary use. According to the author, corpora can be used to study polysemy of words, synonymy as lexical relation, metaphoricity, and idiomaticity of words and register variation: jargon, slang, and appropriateness. Chapter four 'frequency and vocabulary' introduces the importance of frequency in corpus research, the frequency of spoken and written vocabulary, the frequency of function and content words, frequency, and Zipf's law. The author mentions that frequency based analysis helps language teachers to prioritize vocabulary to syllabus design, material writing, and classroom pedagogy. Hence, the author highlights how to use frequency of vocabulary to language pedagogy. The intent of the fifth chapter is to introduce corpora, phraseology and formulaic language. The author discusses corpora, phraseology and formulaic language, and corpus informed lists of phraseology and issues that have a direct impact on the process of identifying and selecting useful vocabulary.

Chapter six covers corpora and teaching vocabulary. The author point out that corpus linguistics influences considerably in language teaching. In the same way, data-driven learning approach (DDL) also helps students to engage in language study and language analysis. Learners use online corpora and concordance through DDL approach (O'Keeffe, McCarthy & Carter, 2007). Language teachers and students can use direct and indirect application of corpus in language classroom according to their preference and purpose. Similarly, the author point out that the corpus is used to prepare textbooks, reference books, and dictionaries indirectly. The learners can also access the corpus data directly through a data-driven approach.

Chapter seven explicates the importance of learner corpora in vocabulary learning. According to the author, learner corpora are the language data produced by L2 learners. These corpora are a large collection of data containing millions of words. In this chapter, the author introduces characteristics of learner corpora and learner

corpora to lexical growth, to language learning, to language assessment, and to material writing. Chapter eight covers specialized corpora and vocabulary. Thus, the application of specialized corpora, register, and genre analysis are elaborated. According to the author, the corpus has played a great role in teaching English for specific purpose. Hence, the author states there is a great linkage between corpus and translation, and the way corpus is applied to teach English for specific purposes and translation.

Chapter nine introduces the idea of discourse, pragmatics, and vocabulary. First, the author defines discourse and discourse analysis. Szudarski also presents the integration of corpus linguistics with discourse analysis. Second, the author demonstrates the use of corpus to study lexical features. It is also mentioned that corpus facilitates pragmatic analysis. Hence, the author discusses how corpus can facilitate the analysis of the pragmatic function of vocabularies. The main intent of chapter ten is to point out future research projects in corpus linguistics. The author mentions a wide range of vocabulary issues that can be conducted as future research projects such as phrasal and single word verbs and morphological complexity. In each research project, the author proposes research questions, background, and methodology.

To sum up, the book is comprehensive and user-friendly reference book; as the author explains, it can be a good resource for students, teachers and language practitioners with little or no experience in corpus analysis. Besides, the author answers some questions that arise in using corpus linguistics in language teaching, language analysis, discourse analysis, pragmatics, and research. The other reason that makes this book interesting is that the author has approached the concepts with vivid case studies and examples which help the readers to apply in language teaching and learning contexts. In addition, the qualitative and quantitative analysis in corpus linguistics is a new concept that makes the book selective. The book also provides explanations for corpus linguistics analysis statistically. Finally, though there are many corpus related publications, the book which is user-friendly and comprehensive is mandatory.

## References

- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching* Cambridge: Cambridge University Press.
- Yoon, H. & Hirvela, A. (2004). ESL Student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257-283.

---

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the Journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).