Corpus Linguistics for Vocabulary: A guide for Research by Paweł Szudarsk. Routledge Publications 2018. 239 pp. ISBN: 978-1-138-18721-4

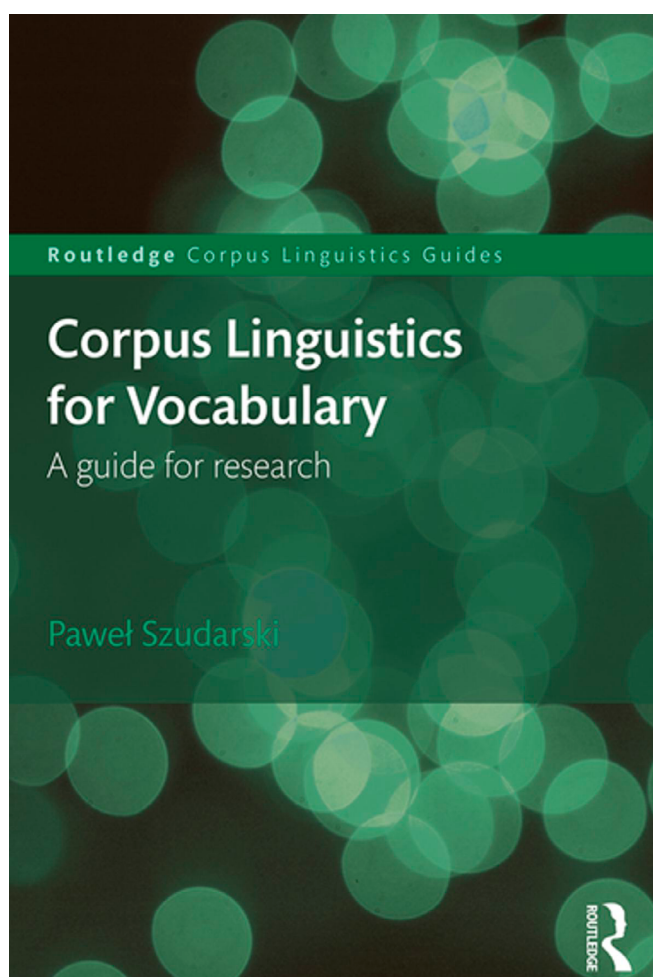**Vahid Pahlevansadegh[1], Mehrdad Vasheghani Farahani[2]**

[1] Georgia University

[2] Universitat Leipzig

Correspondence concerning this article should be addressed to Vahid Pahlevansadegh, Georgia University, 77a, M. Kostava str. Tbilisi, 0171, Georgia. E-mail: vahid_pahlevansadegh@yahoo.com

Since its advent, corpora have been used in various aspects of language-related issues including teaching English (Timmis, 2015). This extensive usage of corpora in language studies has brought about the creation of many resources, among which is "Corpus Linguistics for Vocabulary: A Guide for Research" by Paweł Szudarsk published by Routledge in 2018. This book is composed of 10 chapters. This book provides the competent readership with the systematic and comprehensive clarification of fundamental concepts and terms in corpus linguistics and its applications in vocabulary teaching.

The first chapter on the definition of corpus linguistics comprises five sub-sections. In the first chapter, the author defines corpus linguistics. As the author provides the reader with a very practical and fathomable definition of corpus linguistics, he refers to its main two approaches as corpus-based and corpus-driven; referring to the former as a methodology and the latter as a theory. The second sub-subsection elaborates on the criteria by which a corpus is designed. Having discussed at length and in detail the principles of corpus design and creation, in the third sub-section the author concentrates on the advantages of corpus linguistics by listing the benefits we can take from corpora. In the fourth sub-section, the author sheds light on the limitations of embarking on corpus analysis; pointing to the constraints that corpora can have. In the last sub-section, a variety of internet-based corpus types are provided, especially free databases.

Chapter 2 has three sub-sections and is about tools and statistics in corpus analysis. The first sub-section focuses on tools and types of analysis. First, the author introduces several web-based corpora. Then he uses concordance as an important tool for conducting different kinds of corpus-informed research with some examples. Next, some very important corpus analyses are discussed, such as frequency analysis and concordance, wordlists, cluster (n-gram) analysis, and keyword analysis. In the second sub-section, statistical tests in carrying out corpus-based research are elaborated on. First, log-likelihood is introduced

for conducting comparative analyses on the frequency values between diverse collections of data. Next, tests like t-score and mutual information (MI) test are introduced to determine collocations. Type-token ratio and lexical richness tests are the last tests elucidated on, which are used to examine lexical diversity. The last sub-section explains how to integrate quantitative and qualitative analyses in corpus-informed research.

Chapter 3 has five sub-sections. It elaborates on vocabulary and introduces (as the first sub-section) vocabulary as a core constituent in language use. The second sub-section provides useful terminologies in literature like words, vocabulary, and lexis. This part presents corpus-related terminologies like word form, lexeme, lemma, word family, and lexical item. In next part, the author provides different perspectives by which vocabulary as a multifaceted construct can be approached. The fourth sub-section discusses vocabulary-related notions that can be studied using corpus-based analysis, which consists of the polysemy of words, synonyms, metaphoricity, and register variation. The author, in the last sub-section, provides some research questions pertinent to the issue.

Chapter 4, entitled Frequency and Vocabulary, includes five sub-sections and discusses the importance of frequency in corpus-based linguistics and different types of frequency-based categorization of vocabulary. The first sub-section focuses on the significance of frequency, which is a rudimentary function of corpora. It also elaborates on different types of frequency in vocabulary, such as the frequency of spoken and written words and that of content and function words. Then, the author introduces Zipf's law (1935), which refers to the diverse relationship between the frequency of an occurring word and its rank in the frequency list. Next, the author focuses on the term 'frequency' itself and enumerates different types of word frequency as low-, mid-, and high-frequency words and explains each type. Lexical coverage and bands of vocabulary are two other terms that are covered in this part. In the second sub-section, the author concentrates on frequency-based research on vocabulary load. The third sub-section is about useful vocabulary lists and different examples of such lists and their specifications. In next sub-section, the author considers the impact of corpus-based research on vocabulary and frequency in teaching. Different pedagogical aspects are addressed which are based on frequency and quantitative analysis. In the last sub-section, a study on mid-frequency readers is discussed.

Chapter 5 comprises seven sub-sections and brings the readers' attention to corpora, phraseology, and formulaic language. In the first sub-section, the author discusses how meaning is conveyed beyond single words and through multiword units. It also introduces the theory of lexical priming (Hoey, 2005), which connects corpus-based research with phraseology through a psycholinguistic perspective. In the next sub-section, the author concentrates on formulaic language and defines it as an umbrella term including all types of phraseological units. In the third sub-section, the open-choice principle, the idiom choice principle, and lexico-grammar are elaborated on. In the fourth sub-section, the author describes various types of multiword units like collocation, colligation, semantic preference, and semantic prosody. The fifth sub-section concentrates on another facet of phraseological research, which is register variation referring to diverse types of language used in social contexts. In the sixth sub-section, the author discusses corpus-based lists of phraseology and research that has been carried out in this field in order to determine the most useful phrases and their usage in language. It also focuses on the process of finding useful vocabulary and issues influencing the process directly. A study related to vocabulary lists is discussed in the last sub-section.

Chapter 6 discusses corpora and teaching vocabulary and includes five sub-sections. In the first sub-section, the author introduces two approaches in which corpora are used in language teaching as direct and indirect applications. The second sub-section is concerned with corpora in indirect application and elaborates on how they influence the development of different materials. The next sub-section is concerned with pedagogical corpora and its benefits in the language teaching process. The fourth sub-section deals with teaching oriented corpora, in which the author names the benefits of corpora in language teaching. Finally, in the last sub-section a study about the use of corpora for teaching purpose is provided.

Chapter 7 includes five sub-sections and discusses the significance of leaner corpora when analyzing learner vocabulary. The first sub-section emphasizes the importance of learner language, which is regarded as all forms of language produced by second language (L2) learners represented by a number of peculiar traits. In the second sub-section, characteristics of learner corpora are elaborated on. The third sub-section deals with using learner corpora in language teaching. The author also elucidates on different examples of corpora characterized by the use of English in EFL contexts. At the end, some ways for carrying out research on learner vocabulary

improvement are put forward, such as research based on longitudinal data. The fourth sub-section starts by explaining the ways learner corpora contributes toward having a more valid and reliable assessments of learners' proficiency. There are, according to the book, some challenges and difficulties in using learner corpora for the purpose of language testing, which have been addressed thoroughly in the book. The last sub-section introduces a study on the application of learner corpora for pedagogical purposes.

Chapter 8 elucidates 'specialized corpora' and has six sub-sections. First, the author introduces this type of corpora as an approach to answer some criticism about the use of decontextualized data in corpora. The second sub-section deals with register and genre analysis. It is stated that as different registers exists in different contexts, they possess different syntactical, lexical, and discourse characteristics. Therefore, a huge part of corpus analysis is devoted to the study of register variations by using register-focused corpora. In the next sub-section, the role played by corpora in investigating English for specific purpose (ESP) is highlighted. It is stated that ESP researchers are now interested in investigating different linguistic features in specialized language applications. Next, the author brings attention to the process of identifying ESP vocabulary using corpus-informed approaches. In the fourth sub-section, the profound impact of corpora in investigating English for academic purposes (EAP) is accentuated. Different corpus-informed research focusing on academic vocabulary is presented, including academic word lists as well as phraseology and academic vocabulary lists. In the last sub-section, the author places emphasis on the application of specialized corpora in areas such as translation studies and literacy texts. .

Chapter 9 focuses mainly on studying the use of words through a corpus-informed approach and discourse analysis. It comprises five sub-sections. The first sub-section defines the terms 'discourse' and 'discourse analysis.' Regarding the use of both methods together (corpus-based approach and discourse analysis), the author mentions that although there are basic differences between these two methods, both of them share a concentration on linguistic-related data. In the second sub-section, the author elucidates the interaction between these two approaches, stating that discourse analysis and corpus linguistics are based on qualitative and quantitative analyses. The focus then moves on to lexical features in the third sub-section. In this part, the importance of vocabulary in forming the basic framework of the text and the way through which discourse analysis and corpus-based approach can highlight the significance of vocabulary on the discourse are discussed. Another feature is 'relexicalization', which is regarded as a way in which speakers use one another's vocabulary to paraphrase conversation. 'Intertextuality' is another term in discourse analysis which that investigates how a discourse refers to a prior or future discourse. In this sub-section, a huge number of authentic examples of conversation are presented and researchers can use them in their studies. Corpora and pragmatics is the topic of discussion in the fourth sub-section. First, pragmatics is defined as the study of how meaning is formed and perceived in certain context. Then, its relation to the role of corpora is elaborated on. At the end, some corpus-based pragmatic research is provided, such as the pragmatic analysis of vocabulary as well as semantic prosody and speech acts.

Chapter 10 is a summary and repopulation of the book, which includes some exercises for readers.

Szudarsk's seminal work has significant contribution to the field of vocabulary learning and teaching in second language. In other words, this book provides the competent readership with the systematic and comprehensive clarification of fundamental concepts and terms in corpus linguistics and its applications in vocabulary teaching. Reading each chapter of this impressive book will lay the foundation for further and detailed research in this area of inquiry. This book is useful in that the examples are taken from BNC (British National Corpus) and COCA (Corpus of Contemporary American English) as two main monitor corpora. The so many authentic examples taken from BNC and COCA brings the reader with tangible examples which is one of the strongest points of this book. Moreover, the author has taken direct phots from these two corpora interfaces in order to familiarize the readers with the lay out of the corpora as well as the real examples. The very concise summaries provided at the end of each chapter bring the readers with a recapitulation of the concepts. Another reliable source of information is the list of jargons and terminologies listed at the end of book. By referring to the list, the author will have the chance to read the practical definitions. Moreover, the very positive point of this book is that it gives lots of concrete examples from the two corpora. These examples help the reader to understand the concepts in a practical way.

References

Hoey, M. (2005). Lexical priming: A new theory of words and language. Routledge.
Timmis, I. (2015). Corpus Linguistics for ELT: Research and Practice. Routledge.