

**PASAA**

**Volume 53**

**January - June 2017**

---

---

## **BOOK REVIEW**

---

---

**Supakorn Phoocharoensil**

*Thammasat University, Thailand*

**Doing Corpus Linguistics**

**Authors:** William J. Crawford and Eniko Csomay

**Publisher:** Routledge

**Year of publication:** 2016

**ISBN:** 978-1-138-02461-8

**Pages:** 163 pp.

*Doing Corpus Linguistics* introduces key concepts in corpus linguistics, which is now regarded as one of the most influential linguistic sub-disciplines contributing to research in applied linguistics and TESOL. Known as a reliable tool widely used in authentic language analysis, language corpora are currently integrated into various fields of language study, e.g. lexis, syntax, semantics, pragmatics, discourse analysis, translation, sociolinguistics, language pedagogy, etc., due to the fact that corpus data often surprisingly provide new and interesting knowledge that is unavailable in non-corpus studies or native speakers' intuition (Lindquist, 2009). This book is therefore aimed at familiarizing novice researchers with fundamental technical terms in corpus linguistics. In addition, methods of utilizing already-existing corpora and ways of building researchers' own corpus are presented.

The book is divided into three parts, i.e. PART I Introduction to Doing Corpus Linguistics and Register Analysis, PART II Searches in Available Corpora, and PART III Building Your Own Corpus, Analyzing Your Quantitative Results, and Making Sense of Data. As a convention of textbooks in corpus linguistics, the first chapter defines Corpus

Linguistics, describing major characteristics that make this linguistic subfield necessary for present-day language research (Jones and Waller, 2015). Two common terms, 'corpus-based' and 'corpus-driven' are also clearly illustrated based on Tognini-Bonelli (2001)'s point of view. With the authors' focus on language variation throughout the book, Chapter 2, which is dedicated to register analysis, presents a register analytical framework for corpus data interpretation. Different situational variables are discussed and exemplified, namely email to a friend and email to a boss, news writing and news talk, and classroom versus symposium presentations. A functional interpretation as the next step in a register analysis follows.

Chapter 3, 'Searching a Corpus', is the highlight of PART II. It discusses how researchers can explore the use of lexical items and grammatical constructions in corpus data. The book provides clear instruction on basic searches using a free-of-charge well-known corpus, i.e. Mark Davies' the Corpus of Contemporary American English (COCA), with pictures of concordance lines illustrating the corpus search procedure. Four main areas are introduced in this chapter, i.e. KWIC (Keyword in Context), Collocations, N-grams, and POS Tags. Unlike other corpus linguistic books where readers' background knowledge of N-grams is often assumed, Crawford and Csomay provide a clear definition of the term and state its significance in lexical studies, with adequate instances derived from COCA. Moreover, POS tagging, which is essential to specifying parts of speech of a searched word through COCA, is explained in a comprehensive fashion.

Readers will be grounded with corpus-based research techniques through hands-on experience in Chapter 4 as they are doing series of exercises using publically available corpora, e.g. COCA and COHA. Such exercises are presented in the form of mini research projects in which readers are asked to investigate both vocabulary-based and grammar-based topics. Each of the project-based tasks is conducive to inductive language learning since the introduction to the language problem is transparent to readers, followed by a step-by-step means of corpus search. Those who teach introductory corpus

linguistics will consider these readily-available projects to be useful, in-class assignments to develop students' understanding of carrying out searches and performing a preliminary analysis of the vocabulary items or grammatical patterns being examined.

The third part is devoted to corpus compilation and corpus analysis through relevant statistics. Chapter 5 begins PART III with sample research topics regarding language variation to be further developed into a more complex corpus-based project. The chapter introduces basics in corpus designing and building to serve researchers' own specific purposes. Even those with limited or little knowledge of corpus linguistics should find it easy enough to follow the guidelines given in order to convert texts into a machine-readable format for subsequent data processing through concordance software programs, e.g. AntConc and AntWordProfiler (Anthony, 2014), thanks to several screen-like pictorial examples extracted from those concordance programs.

The next two chapters are closely related in such a way that they both concern statistics for corpus linguists, with more advanced tools introduced in the latter chapter. In particular, Chapter 6 deals with basic terms, concepts, and assumptions corpus linguistic researchers frequently use in their corpus investigation. Examples of these concepts are variable types and functions, variable scales, variable values, and observations. Although fundamental concepts like Quartile, Variance, Standard deviation, are satisfactorily described, explanation of differences between parametric and non-parametric tests in relation to hypothesis testing is rather insufficient, especially for readers with little backgrounds in statistics. Chapter 7 is a continuation of the preceding one. More advanced statistical tests are the central focus of this chapter, in which the authors present difference tests (One-way ANOVA and Two-way ANOVA) and relationship tests (Chi-square and Correlation). More importantly, application of these tests and interpretation of statistical results are available.

Chapter 8 'Doing Corpus Analysis' centers around "the specifics of writing a research paper and presenting research...[the] goal in this

chapter is to provide both students and teachers with some guidelines for how to demonstrate and present their specific research projects.” (Crawford and Csomay, 2016, xii). Examples of situational and linguistic analyses are provided. The chapter concludes with how researchers traditionally report findings from their corpus-based projects. The very last chapter ‘A Way Forward’ encourages novice researchers to pursue their interest in the growing field of corpus linguistics and take into consideration more complex categories of corpus research. The writers also recommend to future researchers certain useful references with regard to register analysis through corpora.

To summarize, *Doing Corpus Linguistics* provides the step-by-step methodology needed for researchers to successfully employ existing language corpora or create their own for linguistic studies. Almost all of the exercises and projects are practical as readers can easily follow the guidelines in search of the target lexical items or syntactic constructions as required. In addition, readers are exposed to a number of practices in corpus data analysis from which they will be able to acquire their corpus linguistic research skills. Despite the authors’ concentration on register or language variation investigation, this book actually proves a valuable handbook for general linguists, graduate students, or TESOL practitioners interested in incorporating a corpus-based approach into their research studies.

### **The Reviewer**

**Supakorn Phoocharoensil** is Assistant Professor at Language Institute of Thammasat University. He is currently teaching various undergraduate courses in English for Specific Purposes, as well as MA. and Ph.D. courses in English Language Teaching. His areas of research interest include second language acquisition, interlanguage syntax and pragmatics, corpus linguistics. He published research articles in a number of international peer-reviewed journals, e.g. *The Journal of Pan-Pacific Association of Applied Linguistics*, *Hong Kong Journal of Applied Linguistics*, and *The Southeast Asian Journal of English Language Studies*.

### References

- Anthony, Laurence. (2014). *AntConc & AntWordProfiler* [Computer program]. <http://www.laurenceanthony.net/software.html>.
- Crawford, W. J. & Csomay E. (2016). *Doing corpus linguistics*. New York: Routledge.
- Jones, C. & Waller, D. (2015). *Corpus linguistics for grammar: A guide for research*. London: Routledge.
- Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.