e-FLT

http://e-flt.nus.edu.sg/

Review of "Doing Corpus Linguistics"

Title	Doing Corpus Linguistics
Authors	William J. Crawford & Eniko Csomay
Year of Publication	2016
ISBN	978-1-138-02461-8
No. of Pages	xii + 163
Place of Publication	New York
Publisher	Routledge

Reviewed by Merhdad Vasheghani Farahani (Leipzig University, Germany)

Corpus is defined as "a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria" (Bowker & Pearson, p. 9). These days, research in the field of Corpus Linguistics is expanding, thanks to the availability of fast and strong computer software. For this reason, every year, new books are being published and more and more research is being carried out in this area of research. Among the many sources produced in this area of inquiry, a very recent one is the book titled "Doing Corpus Linguistics" written by William J. Crawford and Eniko Csomy, two researchers in this field of study. This book consists of 9 chapters.

The first chapter of the book (11 pages) functions as an introductory section that describes the inception of Corpus Linguistics as an area which "is concerned with understanding how people use language in various contexts" (p. 6). This chapter also sheds light on such concepts as corpora and language variations. The authors describe the notion of the corpus at length, citing examples from a very well-known corpus – namely, The Corpus of Contemporary American English (COCA). In addition, they briefly inform the reader of the history of Corpus Linguistics and its various implications. In this chapter, the authors draw the attention of the readers to the field by arguing that any language is composed of rules and systems, which regulates the system of the language. They also provide an outline of the book in the first chapter. Furthermore, they sketch the very basic but necessary principles of Corpus Linguistics.

Chapter two of the book, on register analysis, is composed of 16 pages and of 6 sub-sections. In the first sub-section, focusing on the register, the authors justify the reasons why they chose this notion as the theoretical framework of the book and give different examples to elaborate their claim. In the second sub-section of the chapter, they explain the notion of register analysis, which is the varieties of language that determine the authenticity of language in relation to lexical and grammatical features. In the next sub-section, which is on situational characteristics and identifying variables, they explain the concept of situational variables. In the fourth sub-section of this chapter, the notion of and the reasons for doing functional interpretation are made clear. In the next subsection, the units of analysis are explained and some prior studies in this field of inquiry are discussed. The authors conclude this chapter by giving readers some exercises to gain a better understanding of the notion of corpora and register analysis.

Chapter three of the book (31 pages), on searches in available corpora, is the first practical chapter of the book. In other words, it is in this chapter that the authors introduce some crucially

important concepts in Corpus Linguistics such as *Key Words in the Context* (KWIC), *collocations*, *N-Grams* and *Part of Speech Tags* (POS). These notions are explained with the aid of the COCA. As the authors argue, the reason why this corpus is used in their research is because it is free of charge and is a very large, representative and balanced corpus of contemporary American English. With the illustrations, the reader can easily find his way through the corpus and learn how to use it in the practical context. At the end of this chapter, readers are led to do a search in the COCA. All the steps for doing so are explained. It is not clear, however, why the authors draw their examples from the COCA corpus only, as there are other useful and free-to-use corpora like the British National Corpus (BNC) or Brown Corpus.

Chapter four of this book is on doing projects with corpora which are publicly available. The projects are divided into two main sections as word and phrase-based projects and grammar-based projects. As far as word and phrase-based projects are concerned, the authors embark on five different projects. These examine lexical change over time, the meaning of "literally", "gate" as suffix for any accepted problem after the term "Watergate", clichés, collocations of modifying elements and the word "sustainability". In all of these short projects, the authors try to draw the attention of the readers to the possibilities that corpora can offer in doing register analysis with the variations of words and phrases. On the other hand, the grammar-based projects include six examples. They include grammatical agreement in collective nouns, variations in passive voice, "going to" as a modal verb, grammatical constructions following words such as "begin", "continue" and "start", grammatical categories and coordination, and, lastly. the three words "frugal", "cheap" and "thrifty". The authors give these examples to demonstrate to readers how grammatical rules in English can be interesting topics of research by applying corpora.

Chapter five of the book (10 pages) is another interesting part of the book, because it gives hints and very helpful instructions on how to create one's own corpora (Do It Yourself Corpora.). First, an introduction is given to the notion of the *disposable corpora*. Also, some practical hints are given to the creation of this kind of corpus. Then, the authors give some hints for the creation of a do it yourself corpora. In addition, technical issues are explained as to how to change the format of the raw data into machine-readable texts. The second part of this chapter is dedicated to introducing software and corpus tools. One of the tools introduced is the *AntwordProfler* (Anthony, 2014), with lots of figures of the program so that the reader can easily follow the instructions for using the software and its applications.

It goes without saying that any kind of corpus-based study requires quantitative and statistical analysis (Oakes, 2003). For this reason, the next two chapters of the book are entirely dedicated to statistics for corpus-based studies. In chapter six (15 pages), the notion of the variable and the various kinds of variables are explained. *Central tendency* as a major statistical concept is explained through examples. A very clear distinction is made between *parametric* and *non-parametric* tests as two major statistical tests. In addition, the various types of scales such as *nominal*, *ordinal* and *interval* are delineated. In addition, the concept and types of hypothesis are elaborated on, followed by the distinction between one-tailed and two-tailed tests/hypothesis. In the rest of chapter six, the authors discuss the preparation of raw data and descriptive statistics. In this section, the SPSS program is introduced as one of the most useful statistical software, and through lots of examples and figures, the reader can easily learn to operate the program.

Chapter seven (17 pages) complements the previous chapter by introducing various statistical tests such as ANOVA, Chi-square test, and Pearson Correlation. First, the concept of difference tests, which are tests by which we can see "whether there is a statistically significant difference in the average" (p. 117), is elaborated. One-Way ANOVA and Two-Way ANOVA are explained in detail. Next, the notion of relational tests like the Chi-square test is explained. In addition to the Chi-square test, the Correlation test is also explained. It is not, however, clear why the authors do the analyses manually, while in the previous chapter, they explained, at length, the use of SPSS software for descriptive statistics. This meant that a lot of space is devoted in the book to manual statistical computations.

Chapters eight (10 pages) and nine (4 pages) of the book, the concluding chapters, are dedicated to the conduct of corpus-based research in the real situation. Like chapter one, the book explains the concept of register analysis and its main components. Like other chapters, this chapter contains figures and graphs in order to explain the concept of the register analysis. Particularly useful in this chapter is the part dedicated to the presentation of data in a research paper. It also provides a rubric for researchers to assess the quality of papers.

Chapter nine of the book, which is the shortest chapter, elucidates, very shortly, the future of Corpus Linguistics as a growing area of inquiry which can be used in various fields of language-related studies such as Contrastive Analysis, First and Second Language Acquisition and Translation Studies.

Regarding the pedagogical implications, the book has the shortcoming that it only adheres to one aspect of language study, that is, register analysis. It would be better if the authors could explain other applications of corpora in language-related areas of study, including the Teaching of English as a Foreign Language (learner corpora), Translation Studies (translation and parallel corpora), and Historical Studies (diachronic and synchronic corpora). Also, it would be better if the authors could dedicate a whole chapter to various types of corpora and how one can use them. For example, if a beginning researcher is interested in doing corpus analysis in the field of Translation Studies, this book would be of little help to him.

Moreover, this book does not inform about various internationally well-known corpora that are commercially and/or freely available like British National Corpus (BNC). In addition, the authors could very briefly explain the types of programs and software for corpus-based studies. Nowadays, there are a number of user-friendly and useful software in the field of Corpus Linguistics, some of which will give the user access to particular corpora. An example of these programs is the Sketch Engine software, which is now available for free for many university students.

In conclusion, this book is one of the best sources in the field of Corpus Linguistics for beginners who are highly interested in but have limited knowledge of, this area of research. The strong point of the book is that it provides a step-by-step introduction to the field of Corpus Linguistics in such a way that the potential reader will be able to gain an understanding of the fundamental concepts and rules of doing research in corpus-based research.

References

Anthony, L. (2014). *AntConc & AntwordProfler* [Computer program]. Retrieved from http://www.laurence-anthony.net/software.html

Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. London: Routledge.

Oakes, M. (2003). *Statistics for corpus linguistics*. Edinburgh: Edinburgh Univ. Press.