

Corpus Linguistics for Education- A Guide for Research by Pascual Perez-Paredes: A Book Review

Awatif Alruwaili

Northern Border University

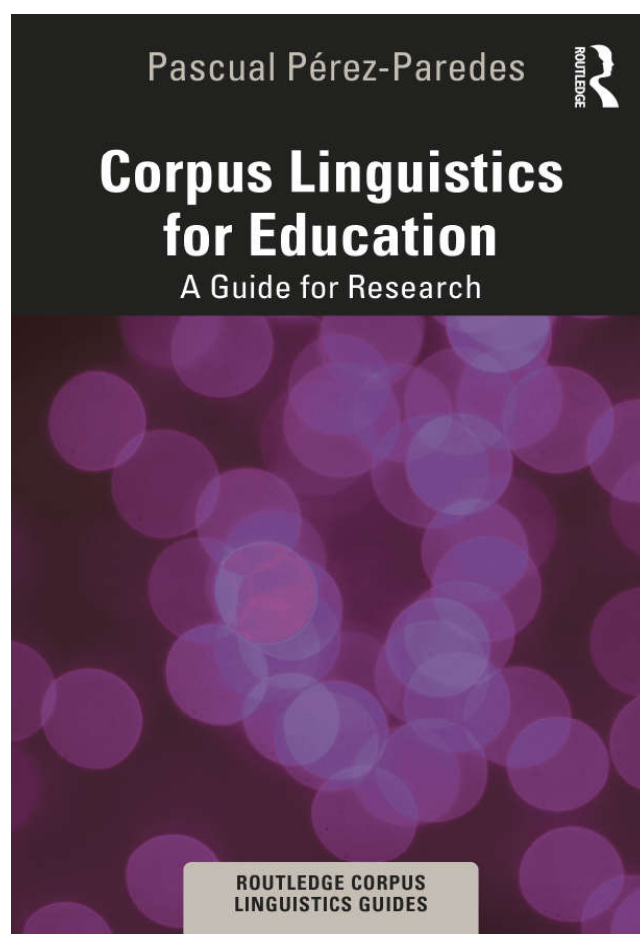
Correspondence concerning this article should be addressed to Awatif Alruwaili, Northern Border University, 9280, Arar Saudi Arabia, Saudi Arabia. E-mail: awatif.k.alruwaili@nbu.edu.sa

Pascual Perez-Paredes, P. (2020), *Corpus Linguistics for Education: A Guide for Research (Routledge Corpus Linguistics Guides)*. Taylor & Francis Ltd.

The book *Corpus linguistics for education* was written by Pascual Perez-Paredes (2020) and published by Routledge. The book aims to provide a guide for how corpus linguistics (CL) methods can be used in educational research. The book consists of seven chapters and an additional conclusion chapter. The book addresses different themes that are relevant for the inclusion of CL methods in the education research field such as frequency register (texts) and keyword analysis, among others. The book is distinguished by the inclusion of a good number of tables and figures providing step-by-step guides for all the selected methods of analysis. This book is very important for researchers and students who are interested in using CL methods in the field of education. This review consists of a brief summary of the eight chapters and a critical discussion of three key issues raised in the book. The review also provides an overall evaluation of the contributions this book has made to this particular discipline.

The first chapter is a thorough introduction to CL and how it can be used in educational research. The author discusses the concept of frequency and gives a number of examples of the use of frequency in learning an L1 and its use in public discourse and texts. This chapter presents two approaches for using CL in education research: the corpus is the primary data and CL is the research methodology, while the second approach uses a variety of methods to collect data that will be analysed by CL methods, so the corpus is considered secondary data. The second approach is the most convenient for use in education research.

The second chapter presents different approaches to text analysis. The author discusses each approach and provides examples of educational research studies using the approach. Content analysis uses a qualitative description approach, which is the 'classical example' (p.21) used in educational research, although it has the lowest level of interpretation. The author describes it thoroughly. In addition, examples of analysis tools such as NVIVO and MAXQDA are provided. The second approach is popular in educational research and has a high level of interpretation. The theme analysis approach is preferably used with a small dataset as it is time consuming because the data will be approached with no presumption (grounded theory). The third approach is conversational analysis, which is used



with spoken discourse. The author mentions the most famous uses of this approach (classroom interaction) as well as its disadvantages, such as being time consuming, a daunting task for analysing every single utterance, full of pauses, etc. The discourse analysis (DA) approach is used for text analysis, a focus different to that of the previous approaches. DA focuses on analysing ideas, stance, power, etc. The author links DA to educational research by stating that both are 'socially committed paradigms that address problems through a range of theoretical perspectives' (p. 25). The last approach is CL analysis of register (texts), which investigates texts from linguistics angles that concern the question of *how*. In a few lines, the author carefully reviews the two directions that respectively see CL as a science (branch of study) and as a mere methodology. The author then explains the types of methods of CL (qualitative by analysing concordance lines and quantitative by analysing collocations and keywords). This chapter attempts to link each approach to an educational research paradigm: grounded theory is a phenomenological qualitative approach, conversational analysis is based on ethnomethodology, discourse analysis is linked to sociocultural theory, and CL methods are linked to different directions taking into account different schools of CL: 1) CL studies texts situated in a positivist research paradigm (Sinclair, 1991) and 2) situated in an objectivism paradigm (Hunston, 2002).

Chapter 3 provides a practical discussion of CL methods and research on language use such as frequency and reading concordance lines. The importance of frequency is discussed, as well as how it helps in discovering language patterns. CL research methods can be used by anyone interested in how language is used in texts. There are many aspects of examining words using CL methods, which expand the areas of inquiry in education research in particular. However, education research may focus on the language as an activity instead of the language as a genre, as linguists may prefer. The author provides two case studies where CL is used as a data analysis method to discover hidden patterns of language use: the first aims to compare the qualitative method (theme analysis) to the CL method (frequency, reading concordance lines). This case study (Fest, 2015) shows how CL methods can complement thematic analysis and go beyond the meanings of words (themes). The second case study uses content analysis and CL methods to analyse language policies of universities in Spain, in which the CL methods can add validity to content analysis findings through statistical testing. The author aims to demonstrate the strength of combining these two methods by selecting these case studies. The author introduces concordance lines by providing a simple definition then moves to the software that is used for reading concordance lines and/or analysing concordance lines (AntConc, Sketch engine). Further, the author introduces practical ways for examining concordance lines while bearing in mind the surrounding context and thinking beyond the word unit. The book introduces the Sinclair model (2003) for reading concordance lines, providing detailed explanation and a step-by-step guide for searching a corpus by using two small corpora.

Chapter 4 introduces CL methods for education research as an alternative for qualitative methods such as grounded theory. The chapter starts by presenting the principles for designing a corpus and how educational research can use the CL methods, either primary or secondary, as a part of the larger methodology of the research. Two case studies were discussed as examples for the use of CL methods as the main research methodology and secondary data as a part of the larger methodology. These two cases investigating education policies may be of interest to the readers of this book. In case study one, the author explains how to design a corpus according to research questions and the steps that needed to be considered before compiling the corpus. Further, it goes in detail through the process for creating and cleaning the corpus. As well as running the analysis (quantitatively and qualitatively), the author helps early-career researchers (postgrads) by showing them an example of breaking research questions into arguments that are linked to corpus building. Case study two uses embedded CL research methods to triangulate the results. The CL method's 'collocation analysis' is used to support the main research methods.

Chapter 5 discusses the interview data within the framework of CL methods and focuses on the process of the interview transcription. This chapter describes interview data and the data that happen during the interview itself. Transcription types are discussed along with the challenges and issues that might come up during the transcriptions. The chapter goes through many details of the transcription process that are rarely found in literature. In addition, an interview transcription can be prepared to be used as corpus data 'building corpus from interview transcription'. The author explains in depth how transcription can be done in a systematic way, reviewing quite a good number of software applications that could help in this process. Moreover, guidelines for orthographic transcription are explained and annotation is covered. The book emphasises the importance of including metadata in the corpus (interview transcription). Text Encoding Initiative (TEI) guidelines are discussed with examples of the metadata included. The book generally is quite good in giving the readers

options in all the discussed methods. For example, it gives the reader two options for annotating a corpus. One is annotation tags in the transcriptions itself, as a quick and easy way for cases where there is limited data. The second option is annotation taxonomy such as TEI, which is more time-consuming but provides a more sophisticated search, especially if the data set is very large. For each option, an example and detailed explanation is presented.

Chapter 6 presents CL methods to investigate the lexis in a corpus (textual data). The first method is keywords, which help to reveal embedded discourse among many other aspects of ideology. The book introduces keywords from a CL research methodology perspective. The author explains the methodological aspects of keywords, multi-keyword analysis, as well as software options used for running such analysis. Step by step, the author takes the reader through building a corpus, 'peace treaties worldwide', running the analysis quantitatively, explaining different types of statistically significant tests such as log likelihood, chi-square, and using reference corpora. The analysis is then continued qualitatively by examining the individual contexts in which keywords occur. Figures and tables are used for further clarification, allowing the reader to follow easily. After identifying the keywords, colligational analysis, another important aspect of CL methods research, is represented, which is an examination of grammatical relationships to discover 'the units of meaning' (p. 133). Noun and noun phrases are used for examining grammatical relationships. Sketch Engine is the recommended software for analysis but it is not the only one mentioned. The author explains another way to examine multi-words, which is n-grams, sequences of groups of words that occur frequently in a corpus. The methodological aspects of n-grams are discussed with examples and software that can extract n-grams from a corpus.

Chapter 7 illustrates a complex corpus search for educational research. It targets analysing talks, a way to integrate CL methods into education research. The author starts by reviewing theoretical aspects of spoken English language analysis from two perspectives: linguistics and educational research. The author explains how to perform a complex search in a very detailed manner with tables and figures. Firstly, a keyword analysis is run, POS tags searched, concordance lines established, and an interpretation drawn. Secondly, lemmas are searched for a specific word and the significant collocation of the searched words is identified. Thirdly, searches for lemmas and symbols are used to customise the findings.

Chapter 8 sums up the book with a reminder of its main aim, which is bridging the gap between CL and educational research. The author acknowledges the challenges that might be encountered in bridging the gap, including the difficulty of collaboration, interdisciplinary work, and the methodological differences between CL and research education. The former is based on positivism and the latter is interpretive paradigms, which have rarely been used to complement each other in published literature.

A Critical Discussion of Three Key Issues Raised in the Book

In this section, I will discuss three issues that the author raised in the book: the collaboration difficulty, education paradigms, and skills presented. The author points out the difficulty of collaboration between CL and education research disciplines because of the differences that might discourage researchers. The main difference is seen as CL being mainly quantitative while education research is mostly qualitative, possibly giving the impression that CL is of no use in educational research. Interdisciplinary work is acknowledged to be difficult in general by the author, and particularly rare between these two disciplines, as the CL discipline is relatively new.

Paradigms in educational research are discussed in one section in the first chapter. Although the explanation is clear and concise, the book could have benefited from the author discussing these paradigms to a greater extent. The author attempts to explain the role of CL methods in education research within the same section but it is not clear enough, and instead of explaining the role, he proceeds to justify why CL methods are not used widely in education research. However, to be fair, the author deals with CL not a science but as a collection of research methods throughout the book (i.e., mere instruments in educational research). Nevertheless, CL paradigms are mentioned in an indirect way in some sections. For example, the author mentions the linguistics-driven paradigm, which is situated in a positivist research paradigm. This paradigm is established by Sinclair, a pioneering corpus linguist (1991). It is worth mentioning that the linguistics-driven paradigm has an approach that is contrary to that of the corpus-based one. This distinction was first defined by Tognini-Bonelli (2001), who defined the linguistics-driven approach as a science and the corpus-based approach as a mere method.

The author covers the most relevant skills that might be needed when using corpus research methods in educational research as well as in the CL field. The book presents the most needed skills for using CL methods as a linguist or as an educational researcher. Eighteen skills were covered and included in the book to discuss the methods. Skill 1, 2, and 7 deal with frequency, its interpretation, and how to handle frequency. Skills 3 and 4 discuss register and textual data. Skill 5 is a practical guide for using an existing corpus. Skill 6 summarises the process of reading concordance lines. Skill 8 helps the reader to understand collocations within the CL perspective. Skill 9 explains corpus design features. Skill 10 explains what is involved in comparing two corpora, while skill 11 explains all the descriptive and inferential statistics mentioned in the book. Skill 12 provides the readers with hints for transcribing data. Skill 13 discusses how to use one's own taxonomy to annotate one's data. Skills 14, 15, and 16 are those related to keyword analysis, researching noun and noun phrases, and using n-grams. Skill 17 explains how to conduct a complex search. Skill 18 is a critical view of CL methods, which is significant for researchers.

A Final Evaluation of the Overall Contribution

This book is very valuable because of its contribution to educational research. CL methods are rarely discussed within the education research framework but this book is devoted to doing so. Each chapter starts without the assumption of prior knowledge by the readers. The book covers most of the important principles in CL such as accountability and how it has a great impact on the findings. The author discusses a very important issue to novice corpus-users, which is the size of the corpus and how it has an impact on the analysis. However, what is more important than the size is the representativeness. Principles of compilation and normalisation of frequency are key for having comparable results, as well as the significant statistical measurement of collocation, which is more valid than native intuitions. The tables and figures facilitate understanding of the CL methods as well as serving as guides to familiarise the reader with applying such methods (analysis). The author is concerned with when to use the CL methods in triangulation, which is critical in interdisciplinary work, so the approach used in this book is "sequential or cyclical" (p. 169). The sequential CL approach is used when CL methods are used as the main component of the data collection method, while the cyclical approach is used when CL methods are used as the starting point for the analysis, then a different method from education research is used, followed by CL methods again in a cyclical process.

References

- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press.
 Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
 Tognini-Bonelli, E. (2001). *Corpus Linguistics at work*. John Benjamins Publishing.