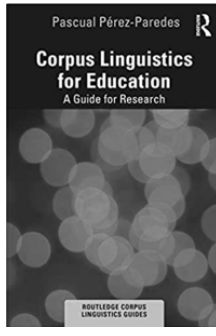


## Corpus Linguistics for Education: A Guide for Research

August 2022 – Volume 26, Number 2

| Corpus Linguistics for Education: A Guide for Research |                                  |               |
|--|----------------------------------|---------------|
| <b>Authors:</b>  | Pascual Pérez-Paredes (2021)     |               |
| <b>Publisher:</b>                                      | Routledge Taylor & Francis Group |               |
| <b>Pages</b>   | <b>ISBN-13</b>                   | <b>Price</b>  |
| Pp. xviii + 175  | ISBN 9780367198435 (Print)       | \$44.95 U.S.. |



Education research, with its large bodies of richly contextualized text, is a field ripe for exploration using corpus methods. In *Corpus Linguistics for Education: A Guide for Research*, Pérez-Paredes provides a practical, entry-level research guide to integrating corpus linguistics in educational research by introducing the ultimate goal of corpus analyses, that is, to generalize empirical description of language use in a target discourse domain through the use of corpora (Egbert et al., 2022). The author shows the potential of corpus methods to improve authenticity and validity regarding traditional textual data and their analyses for better interpretations of research results and findings. Corpus design principles will help reduce bias in the data and allow more inquiries to be answered. Therefore, this book aims to show that corpus linguistics can tap into the research data and that rigorous linguistic feature analyses may help explore education questions. Without assuming specialized knowledge, Pérez-Paredes introduces several key corpus skills for education and entry-level graduate students of any discipline who want to explore corpus methods.

This book is divided into eight chapters, each presenting several skills relevant to educational research, such as why frequency matters, exploring available corpora, and transcribing interview data, among others. Chapter 1 thoroughly introduces concepts of corpus methods and their potential in education. The author emphasizes the role of corpora as samples of language use and why frequencies matter. He then discusses two approaches to using corpora. In the first approach, researchers can use a corpus as their primary or secondary dataset to complement their data resources. In the second approach, researchers can take their existing dataset and analyze it using corpus techniques.

Chapters 2-7 are devoted to corpus-skill building. Chapter 2 provides an overview of several textual analysis methods and their accompanying research paradigms, such as theme analysis with grounded theory and discourse analysis with sociocultural theory. To help clarify that corpus methods also include qualitative analyses besides quantitative (e.g., through frequency counts), the author includes a discussion on register analysis and situational characteristics (see Biber & Conrad, 2009). Chapter 3 covers the most frequently used corpus methods. Fest's (2015) qualitative study of online self-assessment tools is used as an example to illustrate how corpus analysis can be used to complement the qualitative method of interview transcription. In this example, the transcription was part-of-speech tagged and then analyzed using a corpus tool (i.e., AntConc). In Chapter 4, the author describes comparative methods using two corpora, exemplified through case studies on educational policies. Basic design considerations for building corpora are then discussed. Chapter 5 is devoted to the use of corpus methods in education. The great-detailed discussion of corpus structures on interview transcriptions (i.e., metadata and corpus annotations) shows how complex corpus searches can exploit language structure within the transcription, going beyond the exploration of merely transcribed words.

Chapter 6 focuses on the role of vocabulary and its distinct implications for language use. Keyword analyses using peace treaties and children's fiction corpus are illustrated. Built-in statistical analyses within corpus software are discussed to introduce the concept of "keyness" and its implications. Chapter 7 discusses how to run complex searches on spoken language. Drawing on the overarching goal of educational research to investigate how meaning is constructed (Cohen et al., 2018), the author shows that combining lexis (a word) and part-of-speech in a tagged corpus would enable more-detailed linguistic feature searches, allowing for more accurate interpretations.

In Chapter 8, the author provides a critical review of discussions presented throughout the book and acknowledges some challenges of incorporating corpus methods into education research due to disciplinary boundaries and different research paradigms. However, he concludes by arguing that corpus analyses could help data triangulation and validation within education research.

Overall, this book provides a valuable and informative discussion on using corpus methods for education research. All the skills presented in the book are clearly described in a way that is easy to follow. Step-by-step guides to explore each skill are given to help readers to visualize corpus analyses and practice independently. Helpful summaries of the skills and additional notes (e.g., a list of useful existing corpora and corpus software) are presented at the end of each chapter. The detailed discussions throughout the chapters, specifically on data exploration presented in Chapters 5-7, illustrate the distinctive contributions of corpus analyses. The book's demonstration of corpus queries, use of common corpus management tools, chapter summaries, notes for further reading, tables containing terms, figures accompanying the hands-on activities, and chapter-review questions facilitate readers' understanding.

However, this book also comes with some shortcomings. First, in Chapter 1, the author has mentioned two competing research paradigms (i.e., positivism and constructivism) and some qualitative research frameworks (i.e., phenomenology and grounded theory). However, situating corpus linguistics among the many other interpretive frameworks that exist, such as transformative frameworks and pragmatism (see Creswell & Poth, 2018), would help researchers better understand how corpus linguistics relates to their own research perspectives.

Second, while there is a brief discussion of the difference between corpus-based and corpus-driven studies, further discussion of how those two approaches are distinct in their practices and implications would be useful. Additionally, supplementary information about part-of-speech tagging and regular expressions would help readers new to corpus better follow practices presented in Chapters 5-7.

Despite its shortcomings, this book offers fresh insight into incorporating corpus methods within education research, positioning this method as a versatile research methodology that can address the need for more detailed data explorations and lead to better research interpretations. The selected corpus skills are relevant to education research, and the guided practice prompts understanding of the discussions. Therefore, this book is useful for education researchers and entry-level graduate students who wish to expand their research skillset by exploring and analyzing corpora.

### **To Cite this Article**

Lestari, F. (2022), *Corpus linguistics for education: A guide for research*, Pascual Pérez-Paredes (2021). *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 26(2). <https://doi.org/10.55593/ej.26102r1>

### **References**

- Biber, D. & Conrad, S. (2009). *Genre, register, and style*. Cambridge University Press.
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education*. Taylor & Francis.
- Creswell J.W., & Poth, C.N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed). SAGE Publications.
- Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora*. Cambridge University Press.
- Fest, J. (2015). Corpora in the social sciences—How corpus-based approaches can support qualitative interview analyses. *Revista de Lenguas para Fines Específicos*, 21(2), 48–69.

### **About the reviewer**

Febriana Lestari is a Ph.D. student in the Applied Linguistics and Technology program at Iowa State University. Her main research interests center around second language acquisition and corpus linguistics. 0000-0002-6544-2542 <febri@iastate.edu>

Copyright of articles rests with the authors. Please cite TESL-EJ appropriately.