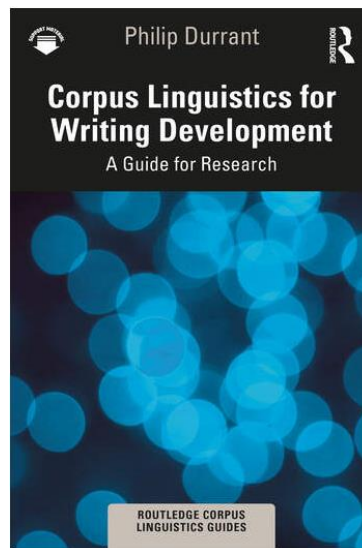## Book Review

"Corpus linguistics for writing development: A guide for research"

İdil Karpuz

MA student, Mehmet Akif Ersoy University, Turkey, karpuzidil@gmail.com

Corpus linguistics for writing development: A guide for research, Philip Durrant. Routledge, (2022). 194 pages. ISBN 9780367715786

Corpus linguistics (CL), as a growing field within the domain of linguistics, is characterized by its systematic analysis of large collections of authentic language data. At its core, corpus linguistics leverages vast corpora—comprehensive collections of written or spoken texts— allowing researchers and educators to discern patterns, frequencies, and nuances in language usage. When integrated into language development, particularly in the realm of writing, corpus linguistics offers a multifaceted toolkit that enriches various aspects of the writing process. Noteworthy in this domain is Philip Durrant's seminal work, ''Corpus Linguistics for Writing Development'' which constitutes a significant contribution to elucidating the pivotal role of corpus linguistics in fostering language acquisition and refinement.

✉**Corresponding Author:** karpuzidil@gmail.com

As a researcher in the field of Applied Linguistics and TESOL, Durrant is an Associate Professor in language education at the University of Exeter. He has conducted considerable research on how corpus tools can improve language teaching and learning, particularly in second-language writing and vocabulary development.

Durrant's book walks the reader through the steps of using corpus linguistics for writing development. His perspective on using corpus linguistics for writing development is beneficial, especially for second-language learners. Throughout the book, the readers are provided with relevant terminology as clearly explained as possible. The explicit advantage of the book could be regarded as its length. It spans one hundred ninety-four pages and successfully maintains the reader's engagement throughout. With a language that is accessible to all, I believe the book could be a departure point for beginners who would like to delve into the details of corpus linguistics and writing instruction. The book consists of four parts and eight chapters. Each part discusses diverge facets of leveraging corpus linguistics for writing development. The first part defines corpus linguistics and CL's fundamental tenets as well as its significance in the context of writing development. The second part focuses on vocabulary in writing development, and part three moves on to studying grammar in writing development. Finally, the fourth part discusses studying formulaic language in writing development. The chapters are organised consecutively. Under each part, there are two chapters in total, eight chapters.

Part one: Foundations

This part aims for an introduction and provides a basis for the reader. It consists of 2 chapters, one and two. The first chapter is named Studying Writing Development with a Corpus. The author starts his words by highlighting the importance of writing. He remarks on writing in L2 learning because it opens doors to international communities and opportunities. He emphasizes the importance by uttering "Writing is an essential social skill that can help people participate in social, educational, professional, and political life. For children learning to navigate the world through their first language, writing is essential for communicating with others and taking part in society. For second and foreign language learners, writing well in their target language can open doors to international communities and increase their life opportunities." (Durrant, 2022, p.3).

Then he dwells on the introduction of corpus linguistics since this approach will be the focus of this book. According to him, corpus linguistics is a method used to examine how individuals employ language. The critical concept involves a researcher who (1) selects a specific area of language usage they find intriguing, such as stories authored by young elementary school students, research reports composed by postgraduate students, or essays written by second language learners for a language proficiency test (2) They gather a set of texts that they consider to be representative of that particular domain, (3) store the collected texts in a digital format, and (4) utilize computer-assisted techniques to analyze and explore how language is utilized within the collected texts. Research done by Klimova (2014) supports this view according to his research, corpus linguistics is a data-driven method of language research that uses large collections of text to study how language is used. Corpus concerns two questions: How does learners' writing change as their education progresses, and what distinguishes more successful learner writing from less successful learner writing. The first question is considered a concern for all educators. The second question, on the other hand, emerges in the context of testing and assessment. The author discusses the corpus's benefits and limitations by providing various aspects. Not only does he provide examples but also offers possible solutions for the limitations. This chapter clearly presents three critical conclusions about research methodology in the context of studying writing development:

(1) Corpus research is an ongoing interpretive process: The author suggests that analysing corpora (extensive collections of written texts) requires continuous examination and interpretation. The patterns researchers observe in their analyses must be constantly questioned and explored from different angles to fully understand their implications for developing writing skills.

(2) Findings should be interpreted broadly: The author emphasizes that differences in language use between learners at different levels cannot be solely attributed to linguistic development. Other factors, such as topic changes, learners' knowledge and attitudes, text or assignment types, and the classroom context, can influence writing patterns. Researchers must be aware of these influences and consider them when interpreting their findings.

(3) Developmental patterns are often specific to local contexts: The author suggests that while some patterns of writing development may be consistent across different learners and learning contexts, many patterns are unique to specific situations. The existing research literature supports this observation, as only a few linguistic variables consistently correlate with development across multiple studies.

Finally, this introductory chapter states the central perspectives this book will handle. It will dwell upon writing development from three standpoints: word use, grammar, and formulaic language. Durrant also emphasizes that vocabulary, formulaic language, and grammar cannot be seen distinctively.

The second chapter is titled 'Learner Corpus Analysis in Practice'. This chapter begins by giving basic information about the tool, that is R. The R Project for Statistical Computing. He advises readers to download it from the website, which is given in this chapter too. The author also provides alternative applications such as R Studio, which might be relatively unknown since there is not much research on the R project. By promoting the tool, the author aims to raise readers' and potential researchers' awareness of such a functional tool. He provides the reason why he selected this tool, "Finally, R has the great benefit of incorporating not only functions that are useful to corpus research but also functions for visualizing and statistically analyzing data. It, therefore, offers a coherent environment in which researchers can interrogate their corpus, visualize the results, and perform statistical analyses." (Durrant, 2023, p.18). This is another reason why this book could be substantially useful for researchers who aim to conduct research on corpus linguistics.

Firstly, the author suggests that readers create a folder in which the examples for this research will be stored. This folder should have four subfolders: corpora, output, R code, and reference. Durrant defines these subcategories, which is excellent if one may need to learn about such details. He continues introducing the program, its importance, benefits, limitations, and how to use it. The author provides the controls and pictures from the app, which is helpful. He also gives examples of the programs he used during the writing process of this book. The steps are in order and explained. They are clear for the reader.

Part two: Studying Vocabulary in Writing Development

This part is related to the vocabulary and consists of two chapters, three and four. The third chapter is 'Understanding Vocabulary in Learner Writing'. The author provides the distinction between breadth and depth of vocabulary knowledge, aspects of word knowledge, and measures of vocabulary development. The fourth chapter, that is 'Vocabulary Research in Practice', presents how different vocabulary analyses can be carried out using R. R is an interesting and rather practical tool for corpus linguistics. At this point, the author suggests that the readers go back to Chapter 2 for file classification and metadata files. The author provides screenshots of

the scripts for a better understanding and visualization. The author starts with R functions and techniques for learner corpus research, moves on with a POS-tagged corpus, and integrates information from an outside reference source for an academic vocabulary corpus. He provides external sources so that readers can easily access them.

Overall, the potential readers could well-equip themselves with how to apply practical techniques to analyze actual learner writing using R for corpus research in this part. These functions and methods are crucial for the other hands-on chapters in the book, making the readers' work here fundamental for the following chapters.

Part three: Studying Grammar in Writing Development

This part is about understanding grammatical constructs within the context of written discourse, encompassing Chapters 5 and 6. Chapter 5 specifically addresses the comprehension of grammar in the context of learner writing, underscoring its intrinsic significance in the realm of writing. Subsequently, the chapter provides a transition to an exploration of methodological approaches, affording the readers a concise introduction conducive to a nuanced understanding of the underlying methodology. In the sixth chapter, Durrant moves on to grammar research in practice. The highlight of this chapter could be the potential to produce misleading results through automated annotation. Durrant endeavours to emphasize this particular facet, emphasizing the intricacies and challenges associated with automated annotation processes.

Part four: Studying Formulaic Language in Writing Development

This last part entails chapter 7 and chapter 8. The seventh chapter dwells on understanding formulaic language. The author explains why he added formulaic language to this book in this chapter. Firstly, he highlights the neglect of formulaic language in research; secondly, since it is a newer area of research in writing development, it needs to be handled separately compared to vocabulary and grammar. As in previous chapters, the author starts with an introduction to the formulaic language. The author defines formulaic language as multiword. These might involve phrasal and prep. verbs, polywords, idioms, proverbs, binomials, pragmatic formulas, collocations, lexical bundles, and lexicalized sentence stems. He provides a table that helps the reader see the bigger picture.

Then, he moves on to how to study formulaic language in the corpus. The chapter offers a step-by-step approach, but it is a start. The actual steps are provided in the eighth chapter. In this chapter, the author imparts the usage of formulaic language in the program. He provides screenshots from the program. According to the author, the most complicated chapter he dealt with was the last. In conclusion, he highlighted the new functions and techniques in this book. Finally, this book delivers valuable insight for researchers interested in corpus linguistics because it offers a tool, useful websites, and step-by-step instructions for using the tool. The author provides examples, templates, and screenshots from the tool. It covers various topics related to corpus linguistics and writing development, such as vocabulary, grammar, formulaic language, academic writing, and language assessment. Compared to other corpus-related books on the market, this book provides more comprehensible input. It explains the theoretical and methodological bases of corpus linguistics for studying writing development. It offers practical guidance on using R for conducting corpus analyses of learner writing, with examples and exercises. It aims at novice researchers and advanced undergraduate and postgraduate students in education, language, and linguistics. The last part (part four) is the most interesting section. Formulaic language is preferred for speaking and listening purposes generally. However, Durrant handles formulaic language from the perspective of writing. After I read this part, I can clearly see that formulaic language is essential for receptive and productive skills.

In sum, Philip Durrant's book *Corpus Linguistics for Writing Development* provides a practical and theoretical exploration of how corpus linguistics can be utilized to enhance writing skills, particularly for second-language learners. Durrant clearly explains essential terminology and methods, using the R programming language to demonstrate practical analyses of language data. The book covers vocabulary, grammar, and formulaic language in depth, presenting real-world applications of corpus techniques in educational settings. It is particularly recommended for novice researchers, as well as advanced undergraduate and postgraduate students in fields related to education, language, and linguistics. Those interested in the systematic analysis of language patterns and their application to writing development will find this book immensely valuable.

Despite its many strengths, the book does present some disadvantages. Firstly, Durrant has not touched upon the effects of a first language on second language writing. This issue, also pointed out by Albelihi (2022) in his review, is significant because the first language background significantly affects the second language learning process. As an educator, I believe including this discussion would have provided a clearer, more comprehensive understanding of the complexities involved. Secondly, the book's relatively high cost could be problematic for some researchers or institutions with limited budgets.

Disclosure Statement

No potential conflict of interest was reported by the authors.

References

Durrant, P. (2022). *Corpus linguistics for writing development: A guide for research* (1st ed.). Routledge. https://doi.org/10.4324/9781003152682

Klimova, B. F. (2014). Using corpus linguistics in the development of writing. *Procedia-Social and Behavioral Sciences*, *141*, 124-128. https://doi.org/10.1016/j.sbspro.2014.05.023

Albelihi, H. (2023). Book review. Durrant, P. (2022). Corpus linguistics for writing development: A guide for research. *System, 115*, 103036. https://doi.org/10.1016/j.system.2023.103036

Copyrights