

Review of Crosthwaite, Peter ed. 2024. *Corpora for Language Learning: Bridging the Research-Practice Divide*. London: Routledge. ISBN: 978-1-032-53722-1. DOI: <https://doi.org/10.4324/9781003413301>

Mohammad Ahmadi  
Lorestan University / Iran

Corpus Linguistics (CL) is an important field of applied linguistics that has enriched the investigation of language in use to a great extent. Today, CL has found application in several areas and has paved the way to new vistas in writing, second language acquisition, lexicography, and related fields. Thus, corpus application is most apparent in the enhancement of writing production since textual analysis of a learner's texts can enlighten such matters as mistakes, tendencies for preferred collocations, or any other questions whenever relevant to a certain topic. Evidently, CL owes significant debt to the work of leading researchers such as John Sinclair (1991) and Susan Hunston (2002), who have revolutionized the quality of language teaching and learning. They have particularly highlighted the fact that data obtained from CL needs to be integrated into language teaching/learning. Their contributions also led to the emergence of specific corpora, carefully tailored to address particular needs of educators, thereby enhancing language teaching and learning (see Römer 2010).

Nevertheless, Peter Crosthwaite's edition *Corpora for Language Learning: Bridging the Research-Practice Divide* has stressed the challenges of incorporating CL into the learning of language. Probably the biggest one is the lack of connection between the latest corpus research and its real-life use in the classroom. This resource is crucial as many educators may be unfamiliar with corpus tools or lack adequate training in data-driven learning (DDL). The volume also elaborates on how the field of corpus research should be integrated into the practices of language teaching. These issues are addressed by providing theoretical concepts and practical guidelines for DDL implementations across



different fields and levels of education, with contributions from renowned international scholars.

The book is structured into 17 chapters, allowing readers to navigate through different sections and follow a coherent sequence of ideas. Initially, Crosthwaite outlines the organization of the book by noting that each chapter is followed by some discussion sections reflecting the perspectives of researchers, teachers or learners who were influenced by that work. The discussions are not in the form of typical research articles; rather, they take the form of personal reflections. In the first chapter, Crosthwaite explains the significance of bridging the research-practice divide in DDL, stressing the need for better connections between academic theories and teaching methods. This is then succeeded by an interview with Laurence Anthony in Chapter 2 where he describes the 2021 update of *AntConc*, which has been redesigned in Python with an SQLite database for proper optimization and improvement. Commenting on the practicality and user-friendliness of *AntConc*, Anthony shows how learners use *AntConc* to do DDL through finding the lexical units by word/keyword lists and n-gram, modifying and distributing new corpora, handling large volumes of data and using effective statistical evaluation. As he notes, “the greatest challenge in DDL for learners is simply finding the target texts and loading them into the concordance” (p. 13).

Chapter 3 is devoted to the use of multimodal corpus data and analytical approaches in language education to improve students’ multisemiotic approaches to meaning-making. Tony Berber Sardinha explains the various ways in which computers can systematically describe images using computer vision techniques, including *Google Cloud Vision Application Programming Interface*. One critical point which is covered in this chapter is the practice of using multidimensional analysis through a particular statistical procedure, known as canonical correlation, which is a corpus linguistics technique, “to detect the dimensions from one particular mode that align with the dimensions from another mode” (p. 28). This allows exploring “discourses, ideologies, and visual content that shape social media conversations” (p. 34).

Chapter 4 is a conversation with Alex Boulton, a DDL expert. According to the text, DDL entails getting students to learn language patterns from corpus data without actually being taught. “They do this not by learning ‘rules’ but by looking at how language is actually used” (p. 43). In the discussion, Boulton specifies the history of DDL, advantages, disadvantages, recent innovations, theoretical implications, technological

advances, and its application to skills beyond writing. He also recommends that greater efforts should be put into the assessment of DDL usability and encourages teachers to introduce DDL activities in the classroom.

The issue of employing DDL methodologies in languages other than English (LOTEs), particularly the L2 context, is the main focus of Chapter 5 authored by Luciana Forti. It can be noted, however, that DDL has been successfully applied for English only, although the author enumerates some possibilities to enhance interaction between DDL, LOTE practices, and SLA theories. The chapter highlights a DDL study on polysemous Italian words and encourages further research to expand DDL application to other LOTE contexts.

In Chapter 6, Ana Frankenberg-Garcia examines a way of applying DDL and corpus tools to enhance learners' appreciation of Academic English. She presents *ColloCaid*, a web-based DDL tool created with academic collocations which enables users to search for collocations, disambiguate words, view concordances and a collocational network, and see example sentences. It is worth noting that the users should not expect the features offered by a professional text editor from *ColloCaid*; rather, it is mainly a "proof-of-concept tool that provides academic English collocation suggestions" (p. 74).

Chapter 7 is devoted to the progression of incidental acquisition through a framework which involves extensive reading (ER) and extensive viewing (EV). Referring to CL findings, Clarence Green stresses the role of comprehensible input in the development of vocabulary and collocations in particular. He further emphasizes the importance of multimedia annotation technology to enhance the comprehensibility of input. Looking at the text in terms of vocabulary difficulty by employing corpus tools, Green recommends that appropriate extensive reading and viewing material should be chosen.

In Chapter 8, Reka R. Jablonkai provides an overview of three main approaches to corpus-based pedagogy, namely, corpus-informed teaching, integrated corpus-supported teaching and learning, and self-directed DDL. Then, she discusses general information about DDL, theoretical background, and a pedagogical model. The chapter offers an insight into various DDL activities, teaching of collocations, lexical phrases as well as discipline specific lexical items.

Chapter 9 sees Tatyana Karpenko-Seccombe looking into the means of adding corpus tools and DDL to enhance the students' argumentation in the academic writing process. The author presents teaching recommendations on corpus consultations, argumentation, patterns of claims and supports, and problem-solution patterns. DDL activities highlighted in the chapter are concerned with such tasks as the concordancers' use to compare collocations, to analyze the distribution of the term across disciplines, or to undertake research with corpora. Focus is given to the tools like *SkELL*, *Lextutor*, and *MICUSP*.

Chapter 10, which is an interview with Tove Larsson and Douglas Biber, cautions against exclusive use of statistical indicators and opaque calculations in CL. It stresses the need for linguistic interpretability and accuracy in research methodology. The chapter further illustrates the peril of relying on quantitative counts without access to annotated texts which can be “problematic if we cannot assess the accuracy of the output” (p.135). The authors also call for a linguistically-motivated paradigm for the analysis of corpora.

Elen Le Foll, in Chapter 11, participates in an interview where she supports open science and education on the interface of CL with language teaching. She especially pays great attention to the issues of openness to knowledge, information sharing, and cooperation. The recommendations include providing free access to research papers, corpus data, and tools, and emphasizes the need for addressing the separation of research and practice in teaching.

In Chapter 12, Agnieszka Leńko-Szymańska has considered how corpora and CL could be applied to assessing learners' L2 vocabulary knowledge. It discusses the issues involved in measuring the extent of word knowledge stressing on the fact that it is not easily measurable since it has many dimensions. The chapter also shows that corpora offer actual language data, and it is possible to develop vocabulary tests on their base. The chapter also discusses the advantages of getting direct access to corpus data and the application of learner corpus data in assessment and modelling of vocabulary.

Chapter 13 is primarily devoted to discussing CL as one of the components of teacher education programs. Qing Ma offers a two-step training framework for the implementation of corpora into teacher education programs for both pre-service and in-service teachers. The challenges and strategies for implementation are described in the chapter, and empirical studies concerning the outcomes of the corpus-based instruction

are outlined. These are followed by the subsequent appeal for more research as to the effects of the range of corpus-based procedures on teacher knowledge and practice.

Chapter 14 focuses on how DDL, CL technology, and phraseography can help improve learners' knowledge of collocation and multiword units. More specifically, Adriane Orenha-Ottaiano presents DDL, describes the activities of DDL on corpora and concordancing tools and focuses on the demand for the accurate frequency of data from corpora. The author also expands the specifics of phraseology and offers corpus-based development activities to be included in the materials.

Chapter 15 ventures into examining broad data-driven learning (BDDL) and its potential when applied to the process of learning informal language supported by technology. Pascual Pérez-Paredes criticizes existing DDL approaches and proposes augmenting DDL to support self-initiated, self-managed learning, utilizing learners' personal electronic data. The author further illustrates examples of tools and resources for BDDL and recommends employing Natural Language Processing and machine learning techniques, as well as, the DDL tasks for the informal settings.

Chapter 16 is concerned with using DDL and CL in the context of EAP when enhancing the internationalization of higher education. Paula Tavares Pinto focuses largely on how DDL and corpus-based activities complement each other, emphasizing their application in analyzing subject-specific academic corpora, teaching academic language patterns, and raising awareness of variations in academic register. The chapter also puts forward the suggestion for the use of the corpora and corpus tools, advantages and limitations of the application of DDL in EAP settings, and the necessity of integrating DDL approaches into the practices in order to assist multilingual scholars and to provide them with the tools to engage in English-medium academic debates.

Chapter 17 of the current volume is specifically devoted to the convergence between CL and EAP. Vander Viana elaborates on the benefits that corpora as well as the corpus-based approaches bring about in EAP research and application. Viana examines how these methods can be used to investigate academic language rigorously, moving beyond intuition or idealized rules. The chapter illustrates methods including compiling specialized academic corpora for the study, genre analysis of texts, cross-comparison of expert and learner writing using learner corpus analysis, and analysis of variations in register across the disciplines using multi-dimensional analysis. Viana further highlights

the significance of enthusiastically produced and selected EAP corpora and also focuses on the interconnection between corpusers (CL researchers) and EAP specialists.

To conclude, the book contains a lot of useful information for the practical implementation of corpus-assisted language learning. However, it could benefit from more critical evaluation and workable solutions for tackling challenges commonly associated with the integration of DDL, and DDL tools into mainstream education. Although the current resource recognizes the problems and potential drawbacks of DDL, it might not sufficiently delve into these issues or provide solutions for addressing them. Moreover, it might not adequately address such issues as pedagogical concerns regarding the relevance of corpus data to specific learning contexts and the potential for these tools to distract from other important aspects of language learning (see Boulton and Cobb 2017).

The book also introduces corpora in vocabulary assessment but fails to discuss other dimensions of the assessment exhaustively. For example, the current coverage could have been improved by articulating how corpora are used in making decisions about grammatical error, pragmatic competence, or specific discursal features. More exemplifications of assessment tasks which are based on the corpus approach, for instance, scrutinizing the texts written by learners with regard to particular linguistic characteristics or devising performance assessments grounded on the actual language data, would contribute to the elaboration of the issue. The expansion of the assessment perspective would offer a better perception of corpora's contribution to the evaluation of language learning results.

Apart from the issues pointed out in the previous section on assessment coverage, this volume also lacks detailed guidance on teacher training in corpus-based language pedagogy (CBLP). While the book advocates the importance of teacher training in CBLP and introduces a two-step framework for developing corpus literacy and pedagogical skills, it did not provide detailed instruction on how to implement it appropriately. In the absence of proper training, the teachers are likely to struggle in implementing some of these novel measures in their teaching programs, leading to suboptimal outcomes for students. For example, in various parts of the book, there are references to the necessity of teacher education. A statement like

CL-literate EAP practitioners are possibly more capable of designing a better endowed exploration of content for discipline-specific EAP, which has implications for teacher education and professional development practices as well (p. 255)

highlights the importance of teacher training but the related sections do not specify a detailed strategy for accomplishing it.

Nonetheless, *Corpora for Language Learning* serves as a useful starting point for those with scholarly interest in CL and DDL. Teachers and scholars in the field of language learning and teaching can learn more about the theoretical framework of the corpus-based language instruction and can get a much deeper insight into the relevant DDL instruments such as *AntConc*, *WordSmith* or *CorpusMate*, as well as the ways for their appropriate use in language learning and teaching. Overall, this resource is very valuable for anyone intending to improve their knowledge in the area of data-driven language education in different settings.

#### REFERENCES

- Boulton, Alex and Tom Cobb. 2017. Corpus use in language learning: A meta-analysis. *Language Learning* 67/2: 348–393.
- Hunston, Susan. 2022. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Römer, Ute. 2010. Using general and specialized corpora in English language teaching: Past, present and future. In Mari Carmen Campoy, Begoña Bellés-Fortuno and María Lluïsa Gea-Valor eds. *Corpus-based Approaches to English Language Teaching*. London: Continuum, 18–35.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

*Reviewed by*  
 Mohammad Ahmadi  
 Department of English Language  
 Lorestan University  
 Khorramabad, Lorestan Province  
 68151-44316  
 Iran  
 e-mail: [ahmadi.m@lu.ac.ir](mailto:ahmadi.m@lu.ac.ir)