

<https://doi.org/10.17323/jle.2025.26958>

Corpora for Language Learning: Bridging the Research-Practice Divide: A Book Review

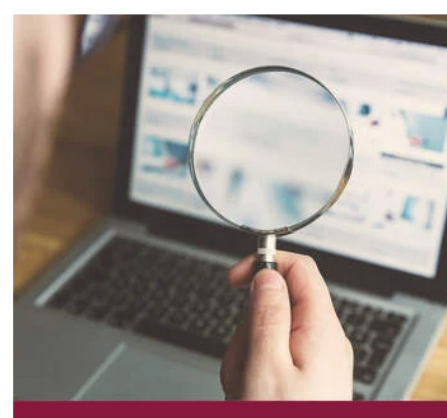
Nguyen Huu Chanh 

University of Health Sciences, Viet Nam National University Ho Chi Minh City

Corpora for language learning: Bridging the research-practice divide by Peter Crosthwaite, New York: Routledge, 2024, 288 Pp., (Paperback), ISBN: 9781032537221

Corpus linguistics has gained considerable attention in applied linguistics over recent decades, particularly for its potential to inform and improve language teaching and learning pedagogy (Conrad, 2005; Stoykova, 2014; Yan & Ma, 2025). However, much of the existing literature tends to focus on theoretical or experimental research, with less attention on how data-driven approaches are systematically integrated into everyday classroom practice. To fill this gap, *Corpora for Language Learning: Bridging the Research-Practice Divide* by Peter Crosthwaite, is committed to examine the challenges of applying corpus-based methods in practical teaching contexts. It provides a detailed account of various corpus tools, methods, and case studies that aim to support language educators, researchers, and practitioners in utilizing corpus linguistics more effectively. The volume contributes to the field by offering concrete examples of how corpus data can inform language acquisition and instruction, especially in vocabulary development, academic writing, and argumentation skills.

This volume provides a detailed examination of corpus-assisted language learning and its practical applications within language pedagogy, with a focus on narrowing the divide between research and classroom implementation. It addresses a broad spectrum of topics, including foundational concepts of Data-Driven Learning, the use of specific corpus analysis tools, and the incorporation of corpus-based methods across diverse educational settings such as teacher training, assessment, and language-spe-



CORPORA FOR LANGUAGE LEARNING BRIDGING THE RESEARCH-PRACTICE DIVIDE

Edited by
PETER CROSTHWAITE



cific contexts. The chapters collectively emphasize the potential of corpus linguistics to inform language teaching practices, while also discussing the roles of technology, open educational resources, and promoting learner autonomy.

Several chapters focus on corpus tools that support Data-Driven Learning (DDL) in classrooms. Laurence Anthony's AntConc remains a foundational, accessible concordancer, while Ana Frankenberg-Garcia's ColloCaid specifically targets academic collocations, a crucial skill for advanced learners. Tatjana Karpenko-Seccombe offers practical corpus-based activities to develop argumentation skills, expanding pedagogical uses of corpora. Tony Berber Sardinha's exploration of multimodal corpora in-

Citation: Huu Chanh, N. (2025). [Review of the book *Corpora for language learning: Bridging the research-practice divide* by P. Crosthwaite]. *Journal of Language and Education*, 11(3), 173-175. <https://doi.org/10.17323/jle.2025.26958>

Correspondence:
Nguyen Huu Chanh,
nguyenuuchanh06@gmail.com

Received: May 08, 2025

Accepted: September 18, 2025

Published: September 30, 2025



introduces innovative perspectives by integrating visual elements, though this approach may require additional classroom scaffolding. Overall, these chapters provide valuable resources but would benefit from deeper discussion on practical classroom integration to fully address both research and practice.

A core theme is the application of corpus linguistics to enhance language competencies. Luciana Forti expands DDL beyond English, addressing an important research gap. Clarence Green highlights corpus-informed curriculum decisions for vocabulary acquisition through extensive reading and viewing, while Reka Jablonkai and Adriane Orenha-Ottaiano focus on corpus-supported teaching and phraseological competence development. Paula Tavares Pinto connects DDL with global education goals, adding a socially relevant dimension. This group showcases strong pedagogical innovation, translating corpus research into actionable teaching strategies, though more attention to learner engagement and challenges would strengthen their practical relevance.

This section addresses the crucial role of educators in adopting corpus-based methods. Qing Ma advocates equipping pre- and in-service teachers with corpus literacy effectively. Agnieszka Leńko-Szymańska offers corpus-based approaches for more rigorous vocabulary assessment. Elen Le Foll emphasizes open science and education to enhance accessibility and sustainability of corpus resources, while Tove Larsson and Douglas Biber call for methodological transparency to ensure linguistic interpretability. These chapters highlight systemic factors essential to bridging the research-teaching gap, combining theory and ethics, though more practical examples would enhance their classroom impact.

The final theme broadens DDL beyond formal settings into informal and lifelong learning contexts. Peter Crosthwaite and Alex Boulton set the stage by emphasizing collaboration and accessibility. Pascual Pérez-Paredes introduces Broad Data-Driven Learning (BDDL), reflecting the rise of digital literacies in informal environments. Vander Viana promotes autonomous corpus use in English for Academic Purposes, acknowledging current limitations but advocating for learner empowerment. These chapters push DDL's boundaries and highlight adaptability as key to closing the research-practice gap, though further guidance on supporting informal learners would strengthen these forward-looking insights.

This volume makes a significant contribution to the field of corpus-assisted language learning by exploring practically and comprehensively how corpus linguistics can be integrated into language teaching. One of its greatest strengths lies in thoroughly applying corpus-based methods to address concrete classroom challenges such as vocabulary acquisition, academic writing, and argumentation skills. The volume excels in showcasing concrete tools, such as AntConc (e.g., Chapter 2 by Laurence Anthony), ColloCaid (e.g.,

Chapter 6 by Ana Frankenberg-Garcia), and Lextutor Concordancer (e.g., Chapter 9 by Tatyana Karpenko-Seccombe), and providing case studies and practical activities that make corpus linguistics accessible and actionable for educators. By linking theory with real-life classroom applications, the volume helps address the often-cited research-practice gap, offering educators tangible ways to enrich their teaching with authentic language data.

Further strength of the volume is its detailed attention to corpus tools and resources that support various aspects of language learning. Chapters dedicated to Data-Driven Learning (DDL) approaches (e.g., Chapter 16 by Paula Tavares Pinto) emphasize not only the functionality of tools such as AntConc for corpus analysis but also how these resources can be embedded into lesson plans to enhance learning outcomes. The ColloCaid tool's role in improving collocational accuracy in academic writing and the use of extensive reading and viewing to support vocabulary growth also demonstrate the volume's capacity to address diverse language skills with corpus technology. Including the practical guidance tailored to different learner profiles and teaching contexts, such as English for Specific Purposes, makes these resources both relevant and versatile for a range of language educators.

Moreover, the volume stands out for its focus on the pedagogical implications of corpus linguistics, particularly its potential to foster learner autonomy (e.g., Chapter 8 by Reka R. Jablonkai) and evidence-based instruction (e.g., Chapter 12 by Agnieszka Leńko-Szymańska). By encouraging learners to engage directly with language patterns and corpus data, the volume advocates for a more active and informed approach to language learning. This learner-centered orientation is well supported by chapters that explore not only tool usage but also instructional design and task development, allowing teachers to create tailored, needs-based lessons. Emphasizing transparency in linguistic research methods and the promotion of open science and education also contribute to building a sustainable framework for ongoing teacher development and learner engagement with corpus linguistics.

Despite these considerable and valuable contributions, the volume has some minor limitations that affect its overall applicability and scholarly rigor. To begin with, some chapters (e.g., Chapter 2 by Laurence Anthony and Chapter 11 by Elen Le Foll) tend to focus primarily on their contexts with reliable technological infrastructure, offering little discussion on how corpus methods can be adapted to diverse teaching environments, especially for the educators with restricted resources or institutional constraints. This narrow contextualization reduces the volume's practical value for many educators working in less technologically equipped or lower-resource settings. Future edition could provide sample activities, scaffolding methods, and frameworks for educators unfamiliar with corpus methods.

Furthermore, while some chapters provide strong empirical support through well-documented case studies, others (e.g., Chapter 4 by Alex Boulton) rely mainly on descriptive accounts without much sufficient data to substantiate their claims, which weakens the evidence base for certain pedagogical recommendations. Hence, future editions could offer more data illustrations and teaching tools that advocates the conversations closer to the educators' lived experiences.

Additionally, although the theoretical depth and pedagogical applicability are addressed throughout the volume, the highly technical characteristics of some chapters (e.g., Chapter 10 by Tove Larsson and Douglas Biber) may alienate teachers without specialized training, and some chapters can benefit from a more rigorous theoretical grounding. Therefore, upcoming editions could consider pedagogical summaries or simplified instances to illustrate how the claims and critiques can translate into daily teaching and material developments. Solving these limitations in the following edition would provide a more balanced, nuanced, and globally relevant resource for corpus-assisted language learning.

Overall, *Corpora for Language Learning: Bridging the Research-Practice Divide* makes a valuable contribution to the ongoing dialogue in corpus linguistics and Data-Driven Learning by explicitly focusing on addressing the persistent gap between research and classroom practice. Unlike earlier foundational works such as Johns (1990), which primarily introduced corpus methods into academic language teaching, or Boulton (2010), who addressed accessibility barriers to DDL through innovative delivery formats, this volume consolidates these insights while expanding their application across diverse teaching contexts and learner needs. Moreover, it responds

to recent calls for deeper theoretical engagement, as advocated by O'Keeffe (2021), by combining empirical case studies with pedagogical reflection and open science principles.

What distinguishes this volume is its practical and comprehensive orientation, providing educators, from novices

to seasoned professionals, practitioners, educators,

and researchers with concrete tools, resources, and ev-

idence-based strategies that integrate meaningful corpus into in everyday teaching. Rather than simply revisiting existing knowledge, it pushes the field forward by emphasizing collaborative, transparent, and sustainable research-practice partnerships and by embracing digital literacies and broader educational agendas such as internationalization and lifelong learning. In doing so, the volume offers a comprehensive resource that not only consolidates established DDL methodologies but also charts new pathways for their evolution within modern language education.

ACKNOWLEDGMENTS

I would like to thank the anonymous reviewers for their thoughtful comments and valuable feedback, which greatly improved the clarity and depth of this review.

DECLARATION OF COMPETING INTEREST

None declared.

REFERENCES

- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534-572. <https://doi.org/10.1111/j.1467-9922.2010.00566.x>
- Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 393-410). Lawrence Erlbaum.
- Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14-34.
- O'keeffe, A. (2021). Data-driven learning—a call for a broader research gaze. *Language Teaching*, 54(2), 259-272. <https://doi.org/10.1017/S0261444820000245>
- Stoykova, V. (2014). Teaching corpus linguistics. *Procedia-Social and Behavioral Sciences*, 143, 437-441. <https://doi.org/10.1016/j.sbspro.2014.07.513>
- Yan, J., & Ma, Q. (2025). Theory-supported corpus pedagogy for ESL pre-service teachers: using Parallel EAP Corpora for language learning. *Journal of China Computer-Assisted Language Learning*. <https://doi.org/10.1515/jccall-2024-0016>