



**Vyatkina, Nina. (2025). *Corpus applications in language teaching and research: The case of data-driven learning of German*. Routledge. Pages: 210. ISBN: 9781032122182.**

NGUYEN HUU-CHANH & THAI KHAC-MINH\*

*University of Health Sciences, Viet Nam National University Ho Chi Minh City (Vietnam)*

Received: 27/06/2025. Accepted: 27/04/2026

Studies on corpus linguistics and data-driven learning (DDL) have been conducted in language teaching research, featuring more evidence-based and learner-centered approaches (Crosthwaite & Baisa, 2023; Lee et al., 2020; Son et al., 2025). Nevertheless, previous studies largely concentrate on English, with little attention to the practice of DDL for other languages. Nina Vyatkina's *Corpus applications in language teaching and research: The case of data-driven learning of German* bridges this gap by offering a clear exploration of how corpus linguistics and DDL methodologies can improve the pedagogy of German as a second language. Bridging both foundational knowledge and real-world practice, the book addresses enduring difficulties, such as the prevalence of English in corpus studies and the relationship between theory and application. By presenting literature reviews, empirical studies, and materials development, the book provides important perspectives necessary for German language instruction and the wider discipline of language teaching and applied linguistics.

---

\***Address for correspondence:** University of Health Sciences, Viet Nam National University Ho Chi Minh City; Administrative Building YA1, Hai Thuong Lan Ong Street, VNU-HCM Urban Area, Dong Hoa Ward, Ho Chi Minh City (Vietnam); email: [nhchanh@uhsvnu.edu.vn](mailto:nhchanh@uhsvnu.edu.vn)

The book is thematically structured into seven chapters, providing an in-depth investigation of corpus linguistics and DDL in German as a second language teaching. The first chapter introduces the history of corpus practices and addresses important difficulties in language teaching and learning. It then discusses corpus use in teaching the German language, followed by empirical research on the efficiency of DDL applications. An insightful discussion of Instructed Second Language Acquisition (ISLA) and DDL studies emphasizes methodological convergences. The book presents a case study on vocabulary learning through corpus tools and discusses the development of Open Educational Resources (OER) for DDL. Finally, it concludes by revisiting the central divides in the field and outlining future directions for research and pedagogy.

Chapter 1 introduces corpora and the evolution of DDL in language teaching, highlighting key challenges such as limited DDL research for Languages Other Than English, a gap between DDL and instructed SLA research, and the research–practice divide. It also outlines emerging solutions and sets the stage for the rest of the book by presenting its scope, structure, and chapter summaries.

Building on this conceptual overview, Chapter 2 reviews indirect and direct corpus applications in L2 German teaching, tracing their origins, often to L2 English contexts. Indirect uses are categorized by material type, while direct applications are organized chronologically and analyzed by teaching methods and activities. The chapter concludes by summarizing available corpus-based resources for German instruction and identifying key gaps in pedagogical support and development.

Extending this review, Chapter 3 synthesizes empirical research on the effectiveness of DDL interventions in L2 German classrooms. Building on the previous chapter, it reviews direct corpus applications, compares findings from L2 German studies with broader DDL research, and offers a methodological synthesis. The chapter identifies trends, limitations, and gaps, concluding with recommendations for future pedagogical and research directions in DDL for German.

While the preceding chapters focus on corpus applications within DDL research, Chapter 4 offers a comparative methodological analysis of DDL and ISLA research on teaching L2 collocations. It addresses the disconnect between the two fields and highlights collocations as a shared focus. Using both manual and automated analyses of research articles, the chapter reveals methodological differences and complementarities, concluding with a call for greater integration and collaboration between DDL and ISLA to enhance research and pedagogical practices.

Following this methodological discussion, Chapter 5 reports on a study comparing computer- and paper-based DDL methods for learning German verb–noun collocations. Both approaches led to significant vocabulary gains, but the computer-based method proved more effective for improving morphological accuracy, crucial for inflectional languages like German. Learners with larger vocabularies benefited most. The study supports independent learning and sustainability and demonstrates the value of integrating ISLA and DDL frameworks into vocabulary instruction.

Moving from empirical findings to pedagogical implementation, Chapter 6 explores the development of OER and Open Educational Practices (OEP) for DDL in German language teaching. It discusses the potential and challenges of using open-access corpora as OER, then introduces Incorporating Corpora, a corpus-based OER designed for English-speaking learners of

German. The chapter details its pedagogical design and pilot study findings, concluding with a call for more OER development, especially for Languages Other Than English.

Finally, Chapter 7 concludes the book by revisiting the three key divides in DDL, language, research domain, and research–practice. It summarizes how the book’s findings address these gaps, particularly in the context of L2 German. The chapter also acknowledges the book’s limitations and proposes future directions for expanding DDL research and practice, especially across languages, pedagogical frameworks, and practical applications.

The book offers a comprehensive and insightful exploration of how corpus linguistics can inform language teaching, specifically focusing on German as a second language. One of the book’s major strengths lies in its thorough and systematic coverage of both theoretical foundations and practical applications. Beginning with a historical overview (Chapter 1), the book carefully maps out the evolution of corpus applications and DDL, effectively setting the stage for deeper inquiry. The narrative investigation of using corpus directly and indirectly in German pedagogy (Chapter 2) expands the knowledge of how corpus materials, primarily established for English, can be used in the L2 German setting.

Another significant contribution of this book is the innovative combination of research paradigms. Chapter 4, featuring ISLA and DDL approaches to teaching collocations, emphasizes a notable gap between these disciplines and suggest an integrated model to improve both studies and educational applications. This interdisciplinary analysis not only addresses foundational issues but provides real-world platforms for more successful language instruction. Additionally, the empirical research on the efficiency of diverse DDL approaches for German verb–noun collocations (Chapter 5) offers strong evidence supporting the key values of corpus-based learning, especially on morphological precision, a main dimension for inflectional languages like German.

The book’s concentration on real-world materials is also significant. Chapter 6’s pilot testing of the Incorporating Corpora Open Educational Resource showcases a future-oriented method to make corpus tools usable for educators and learners. By integrating accessible corpora with usage-oriented learning principles, the book provides a scalable framework that promotes independent learner engagement while decreasing teacher workloads and environmental burdens. This real-world trend advocates for the book’s relevance for language teachers looking to integrate DDL into their classroom contexts.

As the book mostly centers on English-speaking learners of German, there is scant attention paid to learners with other linguistic background or diverse proficiency levels. For instance, Chapter 5 primarily emphasizes advantages for high-intermediate students, but it lacks a detailed investigation into adapting DDL methods for beginners or advanced learners. This omission is noteworthy because contemporary language classrooms are increasingly multilingual and heterogeneous, requiring pedagogical approaches that can accommodate diverse learner profiles. Without considering how corpus-based methods may be adjusted for learners with different linguistic backgrounds or proficiency levels, the broader applicability of the proposed DDL practices may remain restricted. Addressing this issue would further strengthen the book’s contribution to ongoing discussions about the scalability and inclusiveness of corpus-informed language pedagogy.

Moreover, there is a lack of exploration of the technological and real-world difficulties relevant to using computer-based corpus tools in the field of language teaching. Even though Chapters 5 and 6 illustrate the application of digital materials, problems such as digital literacy, resource accessibility, and institutional support in less well-equipped contexts receive scant notice. Bridging these gaps more clearly would offer more comprehensive insights and improve the feasibility of using DDL widely. Despite these limitations, the book contributes greatly to the discipline, providing important perspectives for researchers, educators, and curriculum developers seeking to understand corpus-based language learning.

Overall, the book is a timely learning resource that addresses foundational knowledge and real-world use in corpus-based language teaching. Its insightful exploration and forward-thinking models make it particularly valuable for language researchers, educators, and curriculum designers interested in widening their knowledge of data-driven learning in teaching German as a second language. Furthermore, advanced language learners and graduate students seeking applied linguistics and corpus methodologies can also equip themselves with the book's perspectives and highly relevant materials. The book truly serves as a primary knowledge resource for widening corpus use in different languages and pedagogical contexts, promoting future studies and educational innovation.

## REFERENCES

- Crosthwaite, P., & Baisa, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics*, 3(3), 100066.
- Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the establishment of a data-driven learning model: Role of learner factors in corpus-based second language vocabulary learning. *The Modern Language Journal*, 104(2), 345–362.
- Son, J. B., Ružić, N. K., & Philpott, A. (2025). Artificial intelligence technologies and applications for language learning and teaching. *Journal of China Computer-Assisted Language Learning*, 5(1), 94–112.