

Data-driven Learning In and Out of the Language Classroom

Pérez-Paredes, P., & Boulton, A.

(2025). Cambridge University Press, 102 pages

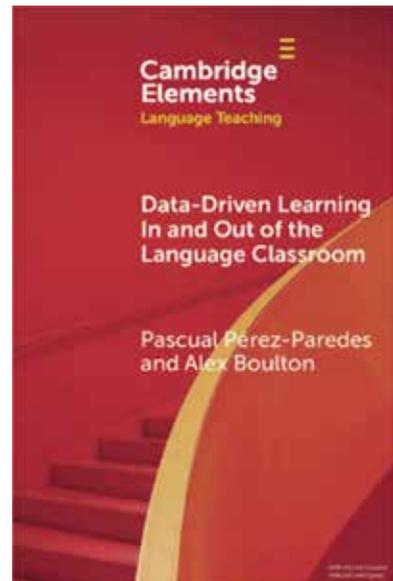
ISBN: 978-1-009-51143-8 (paperback)

JING ZHANG

*School of Educational Studies, Universiti Sains Malaysia*DOI: <https://doi.org/10.30827/portalin.vi45.34881>

Language teachers and learners have access to tons of information every day in the digital world, which demands for extensive explorations about the data-driven learning (DDL), or the use of language corpora for teaching and learning practice (Gee & Hayes, 2011; Crosthwaite et al., 2023; Wang & Derakhshan, 2025). Empirical studies have illustrated the effectiveness of DDL in fostering autonomy and critical thinking in the language learning and teaching practice (Looi et al., 2025). The thriving development of generative artificial intelligence (AI) and large language models (LLMs) further brings about enormous opportunities and challenges of DDL, but there remains a paucity of guidebooks about the effective integration of DDL activities in the new era (Curry et al., 2024). The book, *Data-Driven Learning in and out of the Language Classroom*, edited by Pascual Pérez-Paredes and Alex Boulton, represents timely and significant contributions to illustrating step-by-step procedures of incorporating authentic texts from conventional corpora and LLMs into DDL activities. Overall, this book is well-structured with six chapters by elaborating basic concepts related to DDL, abundant examples of applying authentic language data into different contexts, and cases of generative AI use to foster digital literacy and a critical understanding of language data.

The first chapter opens with basic concepts of DDL and corpus, the historical development of DDL, functions of common general corpora for DDL, alongside recent trends of DDL in the 2020s and beyond. The authors also underscore the critical role of authentic corpora in fostering autonomy, language awareness, and critical thinking. In order to provide readers with an overview of applying corpora into language education, the authors outline a variety of existing corpora (e.g. AntConc, BNC), their applications in DDL, along with elementary, intermediate and advanced DDL activities in Chapter 2. Moreover, this chapter also illustrates the use of concordances that helps learners analyse and identify language patterns in the texts, as well as the basics of querying corpora that teachers can employ into their teaching practice.



Chapter 3 examines the use of pedagogic corpora to develop relevant DDL resources for younger learners. In addition to strategies for creating pedagogic corpora tailored to educational needs, this section provides real-world scenarios from secondary school language teachers, including the process of designing DDL materials for younger language learners, along with the use of language data in the classroom setting. On a different note, Chapter 4 introduces the application of DDL into tertiary education, particularly in teaching academic writing. It elaborates how corpora support language use for both general and specific academic purposes, with detailed examples of building a corpus, writing queries, and interpreting research and results.

The idea of fostering learner autonomy with DDL activities is explored in Chapter 5. By leveraging accessible resources and technology, this section demonstrates how DDL empowers students to take control of their learning process. The authors interpret the use of generative AI tools, such as ChatGPT, as easy-to-query corpus substitutes, and also explain the differences of conventional corpora and LLMs. The concluding chapter synthesizes topics involved in this book and provides forward-looking insights into the future directions of DDL. The authors identify several emerging trends, such as advancements in natural language processing technologies, and highlight the ongoing relevance and necessity of adapting DDL methodologies to align with technological advancements and ever-changing educational paradigms.

This publication provides a timely guide of integrating corpus-based approaches into language education. A notable strength of the book lies in its view of DDL not only as a methodological tool but as a pedagogical philosophy to promote learner autonomy, critical thinking, and digital literacy. Diverse scenarios ranging from pedagogic corpora for young learners to academic writing support at higher education underscore the wide application of DDL activities. Language teachers, graduate students, and professionals in Teaching English to Speakers of Other Languages (TESOL) can draw on this valuable book by recognizing the patterns of authentic language data from corpora for practical use, and following specific procedures to implement DDL activities in language education. Besides, its insistence that DDL activities offer an empirical counterbalance to generative AI tools is of practical relevance because learner competence such as interpretive and analytical skills cannot be replicated by AI tools (Zhao et al., 2025). Therefore, the book is an exceptionally practical resource for teachers and learners to explore the use of corpus linguistics in the language education.

Nonetheless, several limitations warrant our attention. First, the focus on corpora and resources of English language in this book might limit the broad applicability of DDL in multilingual contexts, and overlook the needs of learners with a diversity of linguistic backgrounds. Future studies are recommended to fulfill the needs of acquiring languages other than English (LOTE) with the use of multilingual corpora or generative AI tools. Second, the authors assume that language teachers are provided with sufficient support in terms of digital infrastructure and training to implement DDL activities, which disregard constraints in under-resourced educational environments. To get through such predicament, it is suggested to build open-access digital resources and develop online communities for DDL training, particularly for language learners and teachers in low-resource areas.

Despite these limitations, the book contributes to nuanced discussions of tools like AntConc, Sketch Engine, and broader initiatives like Broad Data-Driven Learning (BDDL),

which demonstrates its commitment to evolving pedagogical ecosystems. In conclusion, this book is an invaluable resource for educators seeking to transform the conventional teacher-centered language learning styles toward a more empirically learner-centred approach. It is particularly commendable for its practical orientation, clarity of presentation, and rigorous integration of corpus linguistics theory with language teaching and learning practice.

REFERENCES

- Curry, N., Baker, P., & Brookes, G. (2024). Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT. *Applied Corpus Linguistics*, 4(1), 100082. <https://doi.org/10.1016/j.acorp.2023.100082>
- Crosthwaite, P., Luciana, & Wijaya, D. (2023). Exploring language teachers' lesson planning for corpus-based language teaching: A focus on developing TPACK for corpora and DDL. *Computer Assisted Language Learning*, 36(7), 1392–1420. <https://doi.org/10.1080/09588221.2021.1995001>
- Gee, J., & Hayes, E. (2011). *Language and learning in the digital age*. Routledge.
- Looi, J., Boulton, A., Hassan, R., & Riget, P. N. (2025). Data-driven learning: A linguistically authentic complement to coursebooks for foreign language learners. *International Journal of Applied Linguistics*, 35, 2148–2171. <https://doi.org/10.1111/ijal.12753>
- Wang, W., & Derakhshan, A. (2025). Exploring the role of EFL students' openness to diversity and self-regulation in their attitudes towards and engagement with data-driven Second Language (L2) Learning. *Learning and Individual Differences*, 121, 102725. <https://doi.org/10.1016/j.lindif.2025.102725>
- Zhao, H., Dang, T. N. Y., & Finlayson, N. (2025). Operationalising critical thinking in postgraduate disciplinary writing: insights from corpus and cluster analyses of lecturer feedback. *Assessment & Evaluation in Higher Education*, 50(3), 422–441. <https://doi.org/10.1080/02602938.2024.2406869>