

## Book review

*Corpus Linguistics and Second Language Acquisition: Perspectives, Issues, and Findings*, by **Xiaofei Lu**. Routledge, 2023, xv+156 pp.

Reviewed by **Shuai Zhang**, Beijing Foreign Studies University, Beijing, China,  
E-mail: zhangshuai2020@bfsu.edu.cn. <https://orcid.org/0000-0002-4170-371X>

<https://doi.org/10.1515/jccall-2024-0001>

## 1 Book introduction

In recent years, there has been a notable surge in the integration of corpus linguistic methods within the domain of second language acquisition (SLA). This trend is evidenced by a growing body of research (e.g., Hunston, 2022; Le Bruyn & Paquot, 2021; McEnery et al., 2019), which has underscored the value of corpus-based methods in renovating traditional SLA research. In this context, the publication *Corpus Linguistics and Second Language Acquisition: Perspectives, Issues, and Findings* by Lu (2023) represents a timely addition to the existing literature. It offers valuable theoretical insights and nuanced methodological guidance for those engaged in corpus-based inquiries of SLA phenomena. This book is poised to serve as an indispensable resource for scholars seeking to advance their understanding of the contributions of corpus linguistics to the field of SLA.

The book comprises six chapters that offer a comprehensive and accessible review and discussion of corpus-based SLA research. Chapter 1 serves as the introduction, defining the primary goal of the publication: to explicitly illustrate how corpus linguistic methods have equipped researchers in the field of SLA with new data sources and analytical techniques. These advancements facilitate the examination of longstanding inquiries and the validation of emerging theoretical propositions within SLA. The term “corpus linguistic methods” primarily refers to “native and learner corpora as data sources, automated and semi-automated tools for corpus annotation, as well as methods, measures, and tools developed for analyzing corpus data in various ways” (Lu, 2023, p. 1). The chapter then articulates four principal issues that the ensuing thematic chapters (Chapters 2 to 5) aim to explore: 1) variation in L2 use; 2) factors influencing L2 processing and production; 3) trajectories of L2 development and influencing factors; and 4) variability and variation in L2 development. Moreover, this chapter introduces theoretical frameworks, notably Complex Dynamic System Theory (CDST) and the usage-based approach, through which these issues are investigated.

Chapter 2 reviews empirical studies that have incorporated corpus linguistic methods to examine variation in L2 use. Lu (2023) first discusses the types of variation in L2 use that have attracted the attention of researchers in the field of SLA, as well as the principal methods employed to measure such variation. Most studies on variation in L2 use have mainly focused on assessing the complexity, accuracy, and fluency (CAF) of the language output, whether spoken or written, generated by L2 learners. Each dimension – linguistic complexity, accuracy, and fluency – has frequently been quantitatively analyzed through an extensive array of measures, such as type-token ratio (TTR) and its various transformations. Subsequently, Lu (2023) systematically reviews the results of empirical studies that used corpus linguistic methods to explore the ways in which L2 use varies due to a range of variables related to the learner and the task. The findings from these empirical studies on the effects of learner- and task-related variables on the variation in L2 use will have “useful implications for verifying the predictions of SLA theories and for informing L2 pedagogy and assessment” (Lu, 2023, p. 39–40).

In Chapter 3, Lu (2023) synthesizes empirical studies that have incorporated corpus analysis to investigate the factors influencing L2 processing and production. He begins with an introduction to usage-based approaches to SLA. Within these approaches, language acquisition is dependent on the linguistic input L2 learners receive and their general cognitive learning mechanisms. Accordingly, the goal of language learning is to acquire constructions, i.e., form-meaning/function pairings. This chapter then reviews the results of empirical studies that used corpus linguistic methods and experimental methods to explore factors affecting L2 learners’ processing and production of certain constructions. The findings from these usage-based studies reveal that L2 learners’ processing and production of constructions are mainly influenced by frequency and contingency. This underscores the need to consider a wide range of constructions and L2 processing tasks to achieve a more comprehensive understanding of the effects of input factors. In addition, Lu (2023) proposes related issues for future studies to consider, including the need for more valid and comprehensive measures to investigate input factors, the necessity of using corpora in highly specialized domains and genres, and the importance of aligning the distributional patterns of target constructions in teaching materials with those in the general language input to benefit L2 pedagogy.

Chapter 4 discusses L2 development trajectories and influential factors. Lu (2023) reviews empirical studies that employed corpus linguistic methods to examine the trajectories of L2 development, as well as the input, learner, and task factors that influence such trajectories. From a theoretical perspective, most studies under review were conducted within the framework of usage-based linguistics (UBL), primarily focusing on the development of L2 in relation to various

types of constructions. However, investigations grounded in alternative theoretical perspectives, as well as those emphasizing different aspects of L2 development – including the development of CAF, as well as the advancement of productive lexical knowledge, phraseological competence, grammatical and morphosyntactic constructions, and academic stance – have also been included. In terms of methodology, these studies adopted a longitudinal design by using longitudinal corpora consisting of language samples collected at different time points from the same learner group to track their L2 development. This chapter concentrates on studies that examine the developmental trajectories of learner groups, instead of including studies focused on one or two individual learners. Notably, Lu (2023) proposes building substantially larger-scale, longer-term longitudinal learner corpora with richer learner and task variables compared to what is presently accessible across various second languages. He also advocates for efforts in the further development of data-driven approaches to dynamically identify and describe L2 learners' development.

While Chapter 4 focuses on research that uses longitudinal corpora to examine group-level trajectories of L2 development, Chapter 5 shifts its attention to individual L2 learners. Specifically, this chapter focuses on reviewing studies concerning inter-learner variability (differences in the longitudinal developmental trajectories among individual learners), intra-learner variability (changes within an individual learner's development trajectory), and individual variation (inter-learner variation in language use or performance at a single moment in time during the developmental process). In this chapter, Lu (2023) also clarifies the major claims and guiding principles of the CDST in corpus-based studies of SLA, which will be beneficial for researchers intending to employ this theoretical stance to guide their future research. By reading this chapter, readers can gain a better understanding of the key assumptions and principles underlying the complex, dynamic systems view of language and language development.

Chapter 6, Lu (2023) first briefly revisits the current perspectives, issues, and findings of the aforementioned four areas of SLA research corresponding to Chapters 2–5, respectively. Then, Lu (2023) summarizes the methodological advantages of corpus linguistic methods for these areas. To promote methodological prosperity in the field, the validity, reliability, and limitations of corpus linguistic methods in facilitating SLA research are discussed. For future research, it is suggested that studies of L2 variation and development could pay more attention to the effects of sociolinguistic variables or variables pertaining to language learning experiences; more attention should be given to specialized and professional domains, compared to the current focus on the general language domain; and greater attention should also be given to the meaning and functional aspects of L2 production.

## 2 Critical evaluation

This book stands as a valuable contribution to the field of SLA. It excels in bridging interdisciplinary research, offers robust methodological insights, and supports the foundation of evidence-based pedagogical practices. Its strengths lie not only in advancing academic discourse but also in practical applications, making it a significant resource for both researchers and educators alike.

Firstly, the integration of corpus linguistics within SLA research facilitates multidisciplinary collaborations, tapping into computational linguistics, psycholinguistics, and sociolinguistics to enhance methodological diversity. This book presents a valuable resource for corpus linguists, illustrating the application of their methods in SLA research and guiding them toward in-depth SLA knowledge. It emphasizes the utility of corpora, which encapsulate real-life language usage, thereby providing SLA researchers with authentic examples of language in various contexts. Moreover, given the expansive availability of large datasets, corpus linguistics enables the examination of extensive linguistic data, surpassing the limitations of manual collection (Zhang & Gu, 2018). Conversely, SLA researchers can remain at the forefront of innovative research while broadening their methodological competence. By promoting interdisciplinary synergy, the book holds the potential to foster mutually beneficial collaborations that will advance both fields in question.

Secondly, the book effectively details the corpus linguistic methodologies applied in contemporary SLA studies, thereby providing a template for future research in the field. It argues that corpus linguistic methods are particularly valuable to SLA scholars due to their capacity to empirically substantiate research findings. As these methods gain traction in the SLA community, they are likely to promote an empirical and positivist shift in the research paradigm. Corpus linguistics empowers SLA researchers to conduct thorough quantitative analyses, identifying patterns and frequencies of language features essential for understanding language learning and use regularities. Beyond the quantifiable, corpus-based research allows for qualitative exploration of pragmatic, discursive, and stylistic language aspects, which are indispensable for a comprehensive understanding of SLA. Moreover, the use of longitudinal corpora provides a dynamic view of language development, mapping out acquisition stages over time (e.g., Ordín & Polyanskaya, 2014).

Last but not least, the book serves as an invaluable asset for emerging teachers and scholars by providing a roadmap for selecting pertinent SLA research topics. By integrating corpus linguistic methods with SLA research, there exists considerable potential to craft language teaching materials underpinned by solid evidence. Corpus

findings can offer concrete guidance for the development of teaching materials and curricula that reflect authentic language use, thereby leading to pedagogical strategies that more effectively meet learner needs (Fang et al., 2021; Ma & Mei, 2021). Moreover, the rich empirical data provided by corpora can stimulate novel research questions and hypotheses in SLA, broadening the field and potentially uncovering new aspects of language learning.

While the book provides a robust examination of the use of corpus linguistic methods in SLA research, its content could be further enhanced by including discussions on how these methods complement traditional SLA research methods, such as questionnaire surveys and classroom observations. Exploring this intersection could offer readers a more holistic insight into the methodological progression in SLA research. Despite this suggestion for improvement, the absence of this discussion does not detract from the book's overall value and applicability.

To conclude, this publication represents an admirable examination of how corpus linguistic methods contribute significantly to advancements in SLA research, thereby serving as an essential reference for scholars undertaking corpus-based SLA studies. It is likely to captivate the attention of both academics and practitioners within the realms of corpus linguistics and SLA, positioning itself as a pivotal resource for those in the intersecting space of these fields.

**Research funding:** This work was funded by the 11th National Foreign Language Education Research Grant in China (Grant No. ZGWYJYJJ11Z043). It was also supported by the Project of Discipline Innovation and Advancement (PODIA)-Foreign Language Education Studies at Beijing Foreign Studies University, Beijing [Grant No. 2020SYLZDXM011].

## References

- Fang, L., Ma, Q., & Yan, J. (2021). The effectiveness of corpus-based training on collocation use in L2 writing for Chinese senior secondary school students. *Journal of China Computer-Assisted Language Learning*, 1(1), 80–109.
- Hunston, S. (2022). *Corpora in applied linguistics*. Cambridge University Press.
- Le Bruyn, B., & Paquot, M. (Eds.). (2021). *Learner corpus research meets second language acquisition*. Cambridge University Press.
- Lu, X. (2023). *Corpus linguistics and second language acquisition: Perspectives, issues, and findings*. Routledge.
- Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 177–190.
- McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus linguistics, learner corpora, and SLA: Employing technology to analyze language use. *Annual Review of Applied Linguistics*, 39, 74–92.
- Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, 42, 244–257.

Zhang, Y., & Gu, Y. (2018). Chinese rhetorical conceptualization of emotion: A corpus-based lexical reconstruction [基于大规模语料库的情感与修辞互动研究]. *Contemporary Rhetoric*, 207(3), 38–54.

## Bionote

### Shuai Zhang

Beijing Foreign Studies University, Beijing, China

[zhangshuai2020@bfsu.edu.cn](mailto:zhangshuai2020@bfsu.edu.cn)

<https://orcid.org/0000-0002-4170-371X>

Shuai Zhang holds a PhD in Applied Linguistics from the School of Foreign Languages and Literature at Beijing Normal University, Beijing, China. He currently works as a teacher and researcher at the Institute of Online Education/Artificial Intelligence and Human Languages Lab, Beijing Foreign Studies University, Beijing, China. His research interests focus on foreign language education, language teacher development, and computer-assisted language learning.