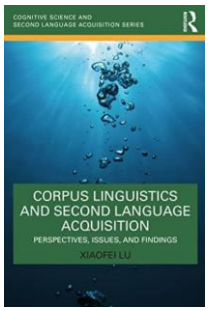
LLT Journal, e-ISSN 2579-9533, p-ISSN 1410-7201, Vol. 27, No. 2, October 2024, pp. 1096-1100

LLT Journal: A Journal on Language and Language Teaching
http://e-journal.usd.ac.id/index.php/LLT
Sanata Dharma University, Yogyakarta, Indonesia

BOOK REVIEW: CORPUS LINGUISTICS AND SECOND LANGUAGE ACQUISITION: PERSPECTIVES, ISSUES, FINDINGS



Title	: Corpus Linguistics and Second Language Acquisition: Perspectives, Issues, Findings
ISBN	: 9781003054948
Author	: Xiaofei Lu
Publisher	: Routledge
Page	: 172

Samsudin Samsudin[1] and Sukarismanti Sukarismanti[2]
[1,2] Universitas Teknologi Sumbawa, Indonesia
samsudin@uts.ac.id[1] and sukarismanti@uts.ac.id[2]
[*]correspondence: samsudin@uts.ac.id
https://doi.org/10.24071/llt.v27i2.7473
received 2 November 2023; accepted 23 July 2024

Abstract
This article discusses the application of corpus linguistic methods in second language acquisition (SLA) research. It highlights the significance of utilizing native and learner corpora, along with automated tools for corpus annotation and analysis, to explore various aspects of SLA. The article is based on a book titled "Corpus Linguistics and Second Language Acquisition: Perspectives, Issues, Findings," which covers five chapters. These chapters delve into topics such as variations in L2 usage at a group level, factors influencing L2 processing and production, group-level longitudinal trajectories of L2 development, and inter- and intra-learner variability in L2 development. The book aims to provide valuable insights and methodologies for researchers interested in employing corpus linguistic methods to enhance their SLA research.

Keywords: corpus linguistics, review, SLA

The corpus linguistic method is a research approach that involves collecting and analyzing large bodies of natural language data known as corpora. This method has proven valuable in second language acquisition (SLA) research, allowing researchers to investigate various linguistic phenomena in a more extensive and authentic context. For example, in this book, the author reviewed a body of SLA research that integrated corpus linguistic methods, including in particular native and/or learner corpora and tools and methods for corpus annotation and analysis, to examine issues in four core research areas, such as learner and task variables that condition variation in second language (L2) use, to understand the effects of various input factors on L2 processing and production, to track group longitudinal

trajectories of L2 development and the input, learner, and task factors that affect such trajectories, and to profile inter- and intra-learner variability and individual variation in L2 longitudinal development.

The book "Corpus Linguistics and Second Language Acquisition: Perspectives, Issues, Findings" covers five chapters. Chapter 1 introduces two key subjects: corpus linguistics meets second language acquisition and an overview of the book. Moreover, in this chapter, the author highlights that the use of the phrase "corpus linguistic methods" encompasses native and learner corpora as data sources, along with automated and semi-automated tools for corpus annotation, as well as various methods, measures, and tools utilized for the analysis of corpus data" (p.1). Using native and learner corpora as data sources, in conjunction with the tools and methods developed for corpus annotation and analysis, has demonstrated considerable efficacy in SLA research, as evidenced by previous studies.

Chapter 2 presents an academic review of empirical studies that have utilized corpus linguistic methods to investigate variations in L2 usage at a group level. The chapter commences by providing a comprehensive overview of the various variations in L2 usage that have been of great significance to SLA researchers, along with the primary approaches employed in measuring such variations. Subsequently, the chapter conducts a thorough and systematic appraisal of the outcomes of previous empirical investigations that have adopted corpus linguistic methods to scrutinize theoretically and pedagogically relevant learner- and task-related factors that affect variations in L2 usage. The use of automated linguistic analysis tools and large-scale learner corpora, representing diverse learner and task types, has significantly broadened the scope and scale of research in this field. Concerning the types of variation in L2 use, studies have predominantly focused on the degree of variance in the complexity, accuracy, and fluency (CAF) of L2 learners' overall production, as well as on their use of particular types of linguistic features or constructions, including morphosyntactic constructions and phraseological units. An extensive set of indices has been developed by researchers for measuring these dimensions of L2 use (e.g.,Lu, 2017). Another area of focus has been on variations in the structural and functional characteristics and usage contexts of L2 learners' use of particular linguistic features or constructions, which are usually analyzed qualitatively.

Then, In terms of learner variables, the existing corpus-based studies of variation in L2 use have heavily concentrated on the effects of L2 proficiency and, to a lesser extent, on the first language (L1) background mainly because these two variables tend to be more readily available in learner corpora than other variables (Gablasova et al., 2022). The investigation of cross-proficiency variation in L2 use has been a central focus of learner corpus studies with a cross-sectional design (e.g., Garner, 2016). In contrast, different proficiency level definitions have been employed in different learner corpora. Increasing efforts have been made to map proficiency levels to a standard scale and improve the comparability and generalizability of findings across various studies. Research in this area has reported diverse patterns of cross-proficiency variation for various dimensions of L2 use and different proficiency ranges, including the absence of cross-proficiency variation in some cases. Using learner corpora that contain language samples produced by learners from diverse L1 backgrounds, learner corpus studies have also

shown that L2 learners' L1 background can be expected to affect many but not all aspects of their L2 use.

Furthermore, typological similarities between the learners' L1 and the target L2 do not always lead to usage patterns that are more similar to those of L1 speakers of the target L2, and it is crucial to consider L1 and proficiency effects on variation in L2 use simultaneously. (e.g.,Shatz, 2019). While the effects of task-related variables on L2 use have been more commonly examined in controlled experimental studies, a small group of learner corpus studies has demonstrated how analyses of learner corpora with language samples representing various modes, genres, task types, and topics could be utilized to explore such effects (e.g.,Yoon & Polio, 2017). These studies have validated the effects of task design features on L2 learners' linguistic performance and have sought to explain these effects from the perspectives of information-processing, functional, or interactive demands.

Chapter 3 presents an academic synthesis of empirical research that has utilized corpus linguistic methods to investigate the factors influencing L2 processing and production. The results of these studies indicate that frequency and contingency have been the most heavily investigated among the input factors under scrutiny in corpus-based investigations. In L2 processing, researchers have used L1 corpora to extract information concerning the frequency and contingency of target constructions. They have examined their impacts on L2 learners' processing of such constructions across various processing tasks (e.g., acceptability judgment and comprehension tasks). In the domain of L2 production, researchers have analyzed learner corpora to obtain characteristics of L2 production of target constructions and examined the association of such features with the frequency and contingency of those constructions in L1 corpora. Overall, significant frequency and contingency effects have been discovered on L2 processing and production of lexical categories, article construction, phraseological units, syntactic constructions, and phonological features. A few usage-based studies have also combined corpus analysis to investigate the impacts of recency, semantic prototypicality, and saliency on L2 processing or production of syntactic or grammatical constructions. Corpus linguistic methods were employed in diverse ways in these investigations. L1 corpora were used to extract instances and usage statistics of target constructions, corpus linguistic tools were utilized to calculate measures of semantic prototypicality, and L2 corpora were analyzed for features of L2 production (e.g., Zhao & Shirai, 2018). These factors were observed to have significant or near-significant effects on L2 processing or production, although their impacts appeared weaker or less ubiquitous than those of frequency and contingency.

Chapter 4 provides an academic review of empirical studies that have employed corpus linguistic methods to track group-level longitudinal trajectories of L2 development and/or investigate the input, learner, and task factors that influence such trajectories. This line of inquiry is largely situated within the theoretical framework of usage-based linguistics. Many studies within this body of literature have employed various corpus analytical tools to investigate longitudinal learner corpora, thereby facilitating an understanding of L2 development in a range of linguistic domains, such as CAF, productive lexical knowledge, phraseological competence, grammatical and morphosyntactic constructions, and stance features (e.g., Yoon & Polio, 2017). Several investigations have explored how relevant

learner and task variables, encoded in the learner corpora, interact with the time spent learning the L2 to influence L2 learners' developmental trajectories. For instance, the studies have examined how learner proficiency and genre impact the developmental process.

Additionally, some studies have examined the effects of different input and linguistic factors on L2 development using longitudinal L1 corpora or large-scale synchronous L1 reference corpora. This line of research has proposed and validated various new measures derived from L1 and L2 corpora to assess L2 development across multiple dimensions. Furthermore, new data-driven approaches have been experimented with to identify areas of significant developmental changes. These studies have contributed substantially to our understanding of the time required for significant developmental changes to occur across different L2 domains, the ways in which these changes materialize, and the interaction between learning time and various input, learners, and task variables.

Chapter 5 has utilized longitudinal L2 corpora to investigate inter- and intra-learner variability and individual variation in L2 development. Studies examining variability over time have been primarily informed by the Complex Dynamic Systems Theory (CDST). This theoretical framework emphasizes the dynamic and adaptive nature of the language system, the nonlinearity of language development, and the interaction among different subsystems involved in the development process. Relevant studies include those conducted by De Bot et al. (2007). Intra-learner variability has been studied through multiple statistical methods, aiming to identify longitudinal patterns of change for individual developmental variables. These studies have also examined the interaction between developmental variables over time and have sought to identify any possible developmental jumps in the developmental process. The results of these studies have provided evidence that intra-learner variability behaves under the principles of dynamic systems. Inter-learner variability, on the other hand, has focused on identifying differences between the developmental trajectories or profiles of two or more individual learners.

In some cases, cohort-level longitudinal or cross-sectional developmental trends have been used as a point of comparison to highlight how individual developmental trajectories may differ from cohort-level developmental trends. Lastly, many studies have utilized mixed-effect models to investigate developmental patterns at the group level while simultaneously quantifying individual variation. For instance, Murakami (2016) utilized such models to study individual variation within L2 learners.

Overall, this book contributes to individuals or researchers with varying levels of knowledge about corpus linguistics and/or SLA theories who have developed a genuine interest in learning about or utilizing corpus linguistic methods in their SLA research. With this target audience in mind, the book strives to present a comprehensive examination of the theoretical perspectives and research findings while explicitly highlighting how corpus linguistic methods have enabled L2 researchers to use new data sources and analyses to address previously unresolved issues and verify new theoretical hypotheses in the field of SL.

References

De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism*, *10*(1), 7–21. https://doi.org/10.1017/S1366728906002732

Gablasova, D., Brezina, V., & McEnery, T. (2022). The trinity lancaster corpus: Development, description and application. *International Journal of Learner Corpus Research*, *5*(2), 126–158. https://doi.org/10.1075/ijlcr.19001.gab

Garner, J. R. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research*, *2*(1), 31–68. https://doi.org/10.1075/ijlcr.2.1.02gar

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, *34*(4), 493–511. https://doi.org/10.1177/0265532217710675

Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*, *66*(4), 834–871. https://doi.org/10.1111/lang.12166

Shatz, I. (2019). How native language and L2 proficiency affect EFL learners' capitalisation abilities: A large-scale corpus study. *Corpora*, *14*(2), 173–202. https://doi.org/10.3366/cor.2019.0168

Yoon, H. J., & Polio, C. (2017). The linguistic development of students of English as a second language in two written genres. *TESOL Quarterly*, *51*(2), 275–301. https://doi.org/10.1002/tesq.296

Zhao, H., & Shirai, Y. (2018). Arabic learners' acquisition of English past tense morphology: Lexical aspect and phonological saliency. *International Journal of Learner Corpus Research*, *4*(2), 253–276. https://doi.org/https://doi.org/10.1075/ijlcr.17006.zha