

---

## Review

---

Ruifan Zhou

(Hong Kong Shue Yan University)

Jianwen LIU

(Hong Kong Shue Yan University)

**Corpus Linguistics for Sociolinguistics: A Guide for Research, Joan O’Sullivan, Carolina P. Amador-Moreno, & Anne Barron. London and New York: Routledge (2025). 282 pp. £31.99, Paperback ISBN: 9781138802049.**

Although corpus linguistics and sociolinguistics share a common ancestry in empirical research, they have long developed along separate trajectories due to methodological and conceptual divides (Baker, 2010). While corpus linguistics prioritizes large-scale text analysis through computational tools (McEnery, 2019), sociolinguistics focuses on how social factors shape language variation (Wardhaugh & Fuller, 2021). This volume, *Corpus Linguistics for Sociolinguistics*, written by Joan O’Sullivan, Carolina P. Amador-Moreno and Anne Barron, responds to this interdisciplinary gap by providing a structured guide to investigating sociolinguistic topics through corpus-based approaches. Readers are equipped to explore language variation across language users, language uses, space and time using corpus linguistic approaches. The book comprises eight chapters, including an Introduction, six thematic chapters and a Conclusion.

The opening chapter establishes a systematic framework for integrating corpus linguistics into sociolinguistic research. It begins by outlining core concepts, approaches and methods from both sociolinguistics—such as its three waves of variation research—and corpus linguistics, including corpus types, two main approaches to analysis (i.e., corpus-based approach and corpus-driven approach) and key analytical methods like frequency counts, collocations and n-grams. The chapter further demonstrates practical applications while addressing challenges such as scarcity and size of spoken corpora.

Chapter 2 examines the interplay between language, gender, and age from a corpus linguistics perspective. It first provides an analysis of gender and language use, exploring how gender-related linguistic practices vary across cultures. The discussion then shifts to the relationship between age variation and language change. The authors introduce widely-used corpora like British National Corpus (BNC) and the Michigan Corpus of Academic Spoken English (MICASE) for empirical investigation. Finally, challenges inherent in corpus-based sociolinguistic research on gender and age are addressed, particularly concerning the availability and nature of socio-contextual information and the need for comparable data across contrasted groups.

Chapter 3 transitions to corpus stylistics. It begins with illustrations of how style markers shape characterization in literary works by authors like Shakespeare and Austen. Expanding beyond traditional literature, the authors analyze authentic speech in fiction through a case study of the intensifier “SO” in contemporary Irish writing. Particularly commendable is the chapter’s forward-looking perspective, which moves beyond literary texts to explore the societal impact of style in

political and environmental discourse.

Chapter 4 probes into the study of geographical language variation through corpus linguistics. It traces the methodological evolution from traditional dialectology to modern sociolinguistics, which is, from dialect atlas compilation to the adoption of sociolinguistic interviews and surveys. A theoretical introduction to major models of World Englishes is followed, including Kachru's three circle model, Schneider's dynamic model, Extra- and Intra-territorial Forces (EIF) model and Mair's world system of Englishes. Then, it practically catalogs a wide array of corpora for empirical research on both native and non-native varieties of English, such as The Corpus of Contemporary American English (COCA) and The Asian Corpus of English (ACE).

Chapter 5 provides a clear and practical guide to the intersection of corpus linguistics and digital communication. It elaborates on two primary methodological approaches, namely web as corpus approach (e.g., WebCorp) and web for corpus approach (e.g., COCA), and discusses major web-based corpora, such as The intelligent Web-based (iWeb) corpus, for investigating language variation and change. Beyond methodology, the authors convincingly argue for the e-language as a distinct variety in registers and explores how this kind of language is shaped by context, social factors and individual agency. Finally, the section on ethical considerations rightly emphasizes that ethical practice cannot be standardized but must be carefully tailored to each project.

Chapter 6 offers a synthesis of corpus linguistics and historical linguistics. It effectively showcases the pivotal role of diachronic corpora—from the pioneering Helsinki Corpus to extensive resources like Corpus of Historical American English (COHA)—in the analysis of linguistic evolution. Then, the chapter practically guides the reader through the know-how to conduct meaningful research in historical sociolinguistics, introducing search techniques while encouraging researchers to use all the data that incorporates complementary resources like Google Books and dictionaries.

Chapter 7 presents a nuanced exploration of the relationship between media language and sociolinguistic change with corpus linguistics tool. Media and language change engage in a reciprocal dynamic: media not only accelerates broader linguistic changes in society, such as shaping language attitudes, but its own discourse simultaneously evolves, trending toward conversationalisation and vernacularisation. Apart from language change, language ideology is also an important aspect of sociolinguistic change. Thus, the authors expound the way media representations shape language ideologies and how corpus methods, such as keyword analysis, can uncover both explicit and implicit ideologies embedded within media discourse.

This volume offers a comprehensive and methodologically rigorous exploration of the intersection between corpus linguistics and sociolinguistics. It systematically demonstrates how corpus tools can illuminate sociolinguistic inquiries—from gender and age-related variation to stylistic analysis, geographical varieties, digital communication, historical change and media discourse. The book's significance lies not only in synthesizing theoretical frameworks and practical applications but also in cataloging diverse corpora and analytical techniques. Although its focus remains primarily on English-language corpora and Western contexts—a limitation that future editions might address through broader linguistic diversity—it represents a substantial contribution to empirically grounded sociolinguistic research.

## References

- Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T. (2019). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Wardhaugh, R., & Fuller, J. M. (2021). *An Introduction to Sociolinguistics*. Chichester: John Wiley & Sons.

**THE AUTHORS' ADDRESSES****First Author****Ruifan Zhou**

MPhil Student

Hong Kong Shue Yan University

10 Wai Tsui Crescent, Braemar Hill, North Point, Hong Kong, CHINA

E-mail: 252701M@hksyu.edu.hk

**Corresponding Author****Jianwen LIU**

Associate Professor

Hong Kong Shue Yan University

10 Wai Tsui Crescent, Braemar Hill, North Point, Hong Kong, CHINA

E-mail: jwliu@hksyu.edu

Received: 18 October 2025

Received in Revised Form: 1 December 2025

Accepted: 23 December 2025