Review: Collins, Luke C. (2019). *Corpus linguistics for online communication: A guide for research.* Routledge.

YUHUA FANG
Universiti Teknologi Malaysia/Wuhan Qingchuan University

RUNDONG ZHAO
China University of Geosciences

Book review

ABSTRACT

EN *Corpus Linguistics for Online Communication: A Guide for Research*, published by Routledge, is a practical handbook for corpus-based online communication research. This volume introduces corpus-based online communication research methods for a variety of research topics with empirical case studies. The volume blends theoretical progress with practical application, making it a useful resource for researchers and scholars in online communication and related fields.

Key words: CORPUS LINGUISTICS, ONLINE COMMUNICATION, APPLIED LINGUISTICS

ES *Corpus Linguistics for Online Communication: A Guide for Research*, publicado por Routledge, es un manual práctico para la investigación de la comunicación en línea basada en corpus. Este volumen presenta métodos de investigación de comunicación en línea basados en corpus para una variedad de temas de investigación con estudios de caso empíricos. El volumen combina progreso teórico con aplicación práctica, lo que lo convierte en un recurso útil para quienes estudian e investigan la comunicación en línea y otros campos relacionados.

Palabras clave: LINGÜÍSTICA DE CORPUS, COMUNICACIÓN EN LÍNEA, LINGÜÍSTICA APLICADA

IT *Corpus Linguistics for Online Communication: A Guide for Research*, pubblicato da Routledge, è un manuale pratico per la ricerca sulla comunicazione online basata su corpora. Il volume presenta metodi di ricerca della comunicazione online basati su corpora, fornendo casi di studio empirici su vari argomenti di ricerca. Il volume combina sviluppo teorico ad applicazioni pratiche, costituendo una risorsa utile per ricercatori e studiosi nel campo della comunicazione online e negli ambiti correlati.

Parole chiave: LINGUISTICA DEI CORPORA, COMUNICAZIONE ONLINE, LINGUISTICA APPLICATA

✉ **Yuhua Fang,** Universiti Teknologi Malaysia/Wuhan Qingchuan University
finn0904@qq.com

Since the introduction of the World Wide Web in the 1990s, online communication as a unique form of language transmission has become a vital part of people's everyday lives. In language studies, corpus approaches enhance the integration of quantitative and qualitative methods, rendering research outcomes more observable and reproducible. However, few internet communication-related corpora exist at present and typically contain obsolete, unrepresentative data, serving as a sub-corpus of the general large-scale and balanced corpus. In the interim, despite the fact that some researchers have acknowledged and utilized corpora in online communication studies, they have confronted the challenge of usually employing the "name" rather than the "nature" of a corpus. The "nature" of corpus linguistics is the study of language in a diachronic or synchronic manner using linguistic theory as a guide, probability and statistics as a method, and language data as the object of study. Nonetheless, a number of existing studies only focus on the "name" of the corpus: they have completed the production of the corpus but have not adopted the appropriate analytical processes to use the produced corpus in several dimensions, thus delivering only simple data suggestions. Moreover, these studies usually interpret their data without the oversight of linguistic research paradigms and fail to achieve a better synthesis of rationalism and empiricism, which reflects the "nature" of corpus linguistics. Therefore, Luke Curtis Collins' *Corpus Linguistics for Online Communication: A Research Guide* fills a void in the field by providing a guide to the application of corpus linguistics to online communication study, ensuring that related research is conducted with a solid theoretical foundation in linguistics and a sound methodology.

Corpus Linguistics for Online Communication: A Guide for Research is part of the Routledge Corpus Linguistics Guides series. The volume offers an instructional and practical guide to conducting research on various forms of online communication using corpus linguistics methodologies by providing practical tasks, and utilizing original data derived from online interactions. The volume is divided into four sections with nine chapters. The first section, comprised of Chapters 1-3, presents the theoretical foundations of corpus linguistics to newcomers to the discipline. Chapter 1 clarifies the meaning of corpus by means of examples from general and special corpora, multilingual corpora, ephemeral corpora, and multimodal corpora. Chapter 2 gives a synopsis of corpus design, construction, and extraction methods. The basics of corpus analysis are presented in Chapter 3. The author describes language features that can be queried using corpus analysis, such as tokenization, N-grams, lemmatization, part-of-speech tagging, collection, semantic categorization, images, and register, before introducing measures for quantifying language features, including frequency, keyness, dispersion, and statistical measures such as T-score, mutual information (MI), chi-squared, log-likelihood (LL), and effect size. For each subject mentioned in Chapter 3, the author provides a clear and extensive definition as well as several examples, making this chapter quite helpful for those who are new to this field.

In Chapter 4, the second section, the author presents three online communication research topics that are compatible with corpus methodology. The first topic is online communication's structural properties: Due to its heteroglossic nature, online communication language has distinct structures stemming from written and spoken language. Taking tweets as an example, using wordlist statistics, Zappavigna (2012, p.27) shows that compared to standard language, Twitter language features more @ symbols, http indications, #topic tags, and RT (retweet) abbreviations than standard language. Therefore, the corpus research approach can be applied to the linguistic structure characteristics of the online communication activities of varied media.

The second topic is the communicative function of online communication. Given that the network facilitates information flow between users, the communicative function is one of the most fundamental and essential aspects of online communication. For this topic, the author especially underlines that *graphicons*, the types of graphical devices most commonly seen in online contexts, which include emoji as well as stickers, GIFs, images and videos, are undermining the function of traditional characters. Thus, corpora can be used with other linguistic theories to examine the contextual structural interconnections or interpersonal functions of graphicons. For example, Skovholt et al. (2014) used a corpus of 1,600 business emails for retrieval and analysis, which pragmatically revealed that in business online communication, the graphic expression :-) is used to express a positive attitude after signing to soften the tone following indicative discourses, and to strengthen the tone following emotional discourses.

The third topic is identity, group, and power in online communication: Extensive liberalization of online communication leads to online disinhibition, which enables scholars to study how individuals manipulate language for self-sculpting or other purposes. In this context, a combination of corpus analysis and sociolinguistics would be beneficial. Hardaker and McGlashan (2016), for instance, collected misogynistic and sexually assaulting comments from social media after a British lady launched a feminist petition, and decided if each commentator represented criminal behavior by calculating its politeness index. As a result, the corpus method can be used to examine antisocial online communication activities such as flaming and trolling.

Chapters 5-8, the third section, demonstrates practical applications of corpus analysis in online communication research, with each of the four chapters containing an empirical case study based on corpus analysis. The research focus of Chapter 5 is internet business communication. The author constructs a corpus consisting of images and texts collected from a company's Facebook page and then investigates the corpus from four perspectives: visuals, non-standard language qualities, keywords, and social actions. Results indicate that the company's Facebook page postings make considerable use of advertising language and regional accents to develop its brand image and attract target customers.

Chapter 6 focuses on online education. The author investigates the acquisition of technical terminology by learners in a massive online open course (MOOC) using the collocation network to illustrate the term "social face." By comparing the collocations used by MOOC learners with those in the British National Corpus (BNC) and Corpus of Contemporary American English (COCA), the author demonstrates that the MOOC learners are still in the early stages of learning and do not have a complete understanding of the technical term.

The research topic of Chapter 7 is network news. The author generates a corpus of articles on the public health topic of super-gonorrhea from the *Daily Mail* and *The Guardian*, then utilizes the UCREL Semantic Analysis System (USAS) to identify and categorize the semantic categories of reader comments. According to the results, the audience groups for the two media are notably different, as are their emphasis on comment focus and comment approach.

The research topic of Chapter 8 is a dating app. The personal data and profiles of Tinder users are gathered to construct a corpus. According to the 3-grams parameter, the combination of three words that regularly appears derived from the corpus, users' descriptions in the "Brief Introduction" column serve three pragmatic functions: self-definition, establishing anticipated relationship parameters, and encouraging others to connect. In this section, the author elaborates on the application of the corpus methodology in online communication research by using case examples covering a broad range of issues. The research techniques and results are described in detail and with clarity, providing a guide for readers to conduct similar research.

The fourth section, Chapter 9, explores the methodological challenges of corpus linguistics in general and suggests ways to apply corpus-based methods to online communication research. The author asserts that future corpus-based online communication research will be characterized by greater reflexivity and transparency, as well as increased emphasis on research ethics. Moreover, as network resources continue to expand, online communication must adapt to fulfill the demand for decentralization, nondiscrimination, bottom-up design, universality, and consensus, which will become crucial study fields for online communication in corpus approaches in the future.

Beginning with theory and proceeding to case analysis, the volume's straightforward and logical organization shows beginners the core technique for doing online communication research in corpus approaches. In addition to listing corpus research fundamentals and essential keywords, this volume presents open access corpora and corpus tools for readers to practice. Moreover, the volume includes reflective questions following each chapter's topic to assist readers to reflect on the chapter's material and broaden their research horizons.

In addition, the author weaves the following two perspectives throughout the volume: first, multimodal corpus analysis plays a crucial role in the study of online communication. Text in online communication is frequently supplemented by multimedia elements such as images and videos for audio-visual supplements, or even the other way around, with text serving as supplemental explanations for multimedia assets. After coding, multimodal corpus analysis helps quantify the relationship between different subjects in online communication and disclose how it affects the audience in this instance. Second, moral ethics must constantly be considered during the corpus gathering process. The author emphasizes that particular themes or types of information within the field of online communication research are frequently extremely sensitive, and that researchers must respect public privacy, seek consent, anonymize, assess potential harm, and complete essential evaluations.

In addition to the numerous positives, it is important to note that the corpus utilized in chapter 8 is not especially representative. The Tinder user profile corpus has fewer than 10,000 words, which may result in ambiguous quantitative results. Even though the size of the corpus is not the most crucial criterion, a corpus must be specified and produced in accordance with the study topic. Before collecting data in Chapter 8, the author sets the diverse genders and sexual orientations of users as variables. However, according to the data obtained by questionnaire, there are no responders of non-binary gender, and heterosexuality and homosexuality are the most prevalent sexual orientations, with fewer bisexuals, pansexuals, and asexuals.

Therefore, to improve the capacity of the corpus and make it more exhaustive, it is important to broaden the scope of the survey questionnaire.

Nevertheless, the limitation does not affect the volume's value. Overall, a corpus provides large and diverse database for online communication research, hence, making it more scientific and empirical. *Corpus Linguistics for Online Communication: A Guide for Research* is a helpful handbook for scholars performing online communication research through the corpus approach. This volume offers a concise and lucid guide to corpus linguistics and online communication study and provides inspiration for potential readers. As a result, they may use corpora in their future studies to produce more driving or interesting results with online data.

References

Hardaker, Claire, & McGlashan, Mark (2016). "Real men don't hate women": Twitter rape threats and group identity. *Journal of Pragmatics*, *91*, 80-93.

Skovholt, Karianne, Gronning, Anette, & Kankaanranta, Anne (2014). The communicative functions of emoticons in workplace emails: :-). *Journal of Computer-Mediated Communication*, *19*(4), 780-797.

Zappavigna, Michele (2012). *Discourses of Twitter and social media*. Continuum.

Yuhua Fang, Universiti Teknologi Malaysia/Wuhan Qingchuan University
finn0904@qq.com

EN	**Yuhua Fang** is a PhD candidate at Universiti Teknologi Malaysia and lecturer at Wuhan Qingchuan University. His research interests include corpus linguistics and translation studies.
ES	**Yuhua Fang** realiza el doctorado en la Universiti Teknologi Malaysia, y es docente en la Wuhan Qingchuan University. Sus intereses de investigación incluyen la lingüística de corpus y los estudios de traducción.
IT	**Yuhua Fang** è iscritto al corso di dottorato presso la Universiti Teknologi Malaysia e docente alla Wuhan Qingchuan University. I suoi interessi di ricerca comprendono la linguistica dei corpora e gli studi traduttivi.

Rundong Zhao, China University of Geosciences
siphon@cug.edu.cn

EN	**Rundong Zhao** is a postgraduate student at the China University of Geosciences (Wuhan). His research interests include corpus linguistics and translation studies.
ES	**Rundong Zhao** es estudiante de posgrado en la China University of Geosciences (Wuhan). Sus intereses de investigación incluyen la lingüística de corpus y los estudios de traducción.
IT	**Rundong Zhao** frequenta un corso di master presso la China University of Geosciences (Wuhan). I suoi interessi di ricerca comprendono la linguistica dei corpora e gli studi traduttivi.