DE GRUYTER

Book Review

Book Review on *Corpus Linguistics and Translation Tools for Digital Humanities: Research Methods and Applications*, by **Stefania M. Maci**, **Michele Sala**, Bloomsbury, 2022, xiv+249 pp.

Reviewed by **Fulu Liang,** College of Foreign Languages, Nankai University, Tianjin, China, E-mail: liang_fulu@126.com. https://orcid.org/0009-0005-1522-0342

Keywords: corpus linguistics; digital humanities; methods; tools; translation studies

1 General introduction

As digital humanities (DH) gradually moves from the niche to the mainstream, its impact has been felt by an increasing number of disciplines in the humanities – including corpus linguistics and corpus-based translation studies. Although both DH and corpus linguistics or corpus-based translation studies involve the use of computers, they have developed independently of each other with little interaction until after 2010. The second decade of the 21st century witnessed the boom of disruptive technologies such as artificial intelligence, big data, cloud computing, blockchain, and virtual reality, resulting in heightened awareness of applying computer technologies to humanities research (e.g., Zheng et al., 2022). This transformation offered an impetus to DH, which then flourished worldwide. Many disciplines hastened to embrace it with a view to borrowing computational methods from DH to foster interdisciplinary activities, improve the digital literacy and data literacy of the humanities, or enhance computational thinking in the field. Arguably, DH will shed light on the latter two and be reinforced in return. However, there is little consensus on the best practices of DH-informed corpus linguistics and corpus-based translation studies that can help us clarify where we are, where to go, and how to go. Fortunately, *Corpus Linguistics and Translation Tools for Digital Humanities: Research Methods and Applications*, edited by Stefania M. Maci and Michele Sala, was published at an opportune time and will hopefully lay the foundation on which future research can be based.

2 Book introduction

This book brings together the three strands DH, corpus linguistics, and corpus-based translation studies. It mainly comprises case studies of various research topics from

a variety of research fields. In 10 chapters, this book begins with an introductory chapter (Chapter 1), which is followed by Part 1 (Chapters 2–5), which focuses on corpus linguistics and DH, and Part 2 (Chapters 6–10), which focuses on corpus-based translation studies and DH.

In Chapter 1, the editors justify their reasons for choosing such a theme for the book by viewing corpus linguistics and DH as being in a part–whole relationship after a critical review of their differences and similarities. According to the editors, "DH is the overarching term for the macro-area of research which analyses texts" (p. 3), while corpus linguistics refers precisely to the 'plethora of methods' mentioned in DH, namely, "the set of principled approaches and tools" (p. 3). The editors then provide a brief introduction to Part 1 (the connection between DH and corpora) and Part 2 (the connection between corpora and translation studies). The rest of this chapter is devoted to briefing readers on the content of each chapter in order to enable them to better follow the thoughts of the authors.

Furthering the discussion, Chapter 2 by Paola Catenaccio discusses two main strands of DH – i.e., (i) the study of computer-mediated communication (CMS) in its various forms and (ii) the use of computer-based techniques for text analysis. It highlights that the traditional theories of CMS do not fully account for emergent technology-derived issues, such as multimodality and multisemiotics. Therefore, Catenaccio puts forward an "adaptive theory approach" to DH, which means that theory development in DH should be adaptive not only to capture the evolution of the object of analysis but also in the sense that it must rely on evidence emerging from corpus-driven (or data-driven) investigation.

In Chapter 3, Marina Bondi demonstrates the use of corpora in cross-cultural genre studies with a case study of Corporate Social Responsibility (CSR) reports. The author first lays the foundation for further discussion by necessitating the integration of lexical categories with semantic and functional, pragmatic perspectives and the employment of corpus linguistics. After a critical review of the cross-cultural analysis of CSR reports aiming to elicit the research question and determine the type of corpora (in this case, full corpus and comparable subcorpora) to be adopted, Bondi reports on the language, size, representativeness, source, and comparability of the corpus. For the detailed analysis, a top-down lexico-grammatical analysis of the generic structure of CSR reports is adopted, followed by a bottom-up semantic and pragmatic analysis using keywords and concordance.

In Chapter 4, Miguel Fuster-Márquez discusses the application of corpus to the extraction and operationalization of lexical bundles (LBs), which broke grounds with the compilation of Biber et al.'s *Longman Grammar of Spoken and Written English* (1999). Fuster-Márquez distinguishes between the phraseological approach and the probabilistic approach in studies on LBs and then focuses on the latter. He highlights that the probabilistic approach is an inductive bottom-up approach to

the identification of LBs, which relies entirely on corpus techniques. Further, the core features for the identification of LBs are reported. Fuster-Márquez ends the chapter with a discussion of bundle size, frequency threshold, and dispersion, which are shared criteria for the two main operational approaches to LB identification: frequency-defined bundles and association-defined bundles.

Furthermore, Chapter 5 by Stefania M. Maci investigates the dissemination of the ketogenic diet (KD) discourse on Twitter. After reviewing related literature on the KD, Maci provides a detailed description of the methodology. The data generated during a designated period of time were collected by searching for keywords and hashtags using Social Bearing, a free Twitter analytics application. Then, quantitative-based analysis was performed on the data with Sketch Engine to identify typical linguistic characteristics. To triangulate the data, another quantitative-based analysis was performed on the data with WMatrix 4 to determine the semantic domains. The research findings exhibit Twitter users' understanding of and attitudes towards KD, shedding light on the dissemination of e-health discourse on digital platforms.

Moving on to corpus linguistics and translation studies, Chapter 6 by Patrizia Anesa examines the use of digital corpora for professional legal translation from the perspective of DH. To begin with, Anesa extends DH from academics to the professional setting of legal translation by situating the area of practice between legilinguistics, translation studies, and corpus linguistics within the overarching concept of DH. She then overviews some existing legal corpora employed in legilinguistics and legal translation and offers a glimpse into the evolving relationship between corpora and specialized legal translation. At the conclusion of this chapter, she discusses in detail the use of corpus in translation tools and processes, translation practice, and translator training.

In Chapter 7, Cinzia Spinzi and Anouska Zummo present a comparative study of emotive language in English and Italian migrant narratives to assess the intention and effect of linguistic choices. To be specific, they adopted the Appraisal Theory and focus on its Affect dimension, which comprises five semantic domains for emotions: un/happiness, in/security, dis/satisfaction, surprise, and dis/inclination. After acknowledging the contributions of DH to the availability of data and software, among others, Spinzi and Zummo report on the data collection from digital museums and the design of the corpus. The interrogation of the corpus was conducted with AntConc (version 3.5.9), focusing on the polarity and strategy of emotive language. The conclusion was drawn by comparing and interpreting the results from the English subcorpus and the Italian subcorpus.

Then, Chapter 8 by Francesca Bianchi et al. introduces us to the set of terminology management affordances with built-in learning analytics for interpreter training. Bianchi et al. first identify the needs of a glossary tool linked to monitoring

and self-monitoring tools (for teachers and students, respectively) and possibly supported by learning analytics technologies. Then, they provide an overview of the existing tools supporting terminology management. According to Bianchi et al., the affordances tailored to their needs comprise a glossary tool, web search tracking and logging functions, and a learning analytics system. Bianchi et al. then demonstrated the performance of the affordances in interpreter training at the University of Salento. They end the chapter with insights into the possible uses of the affordances in the future.

Chapter 9 by Gianmarco Vignozzi applies corpus linguistics to the analysis of the construction and translation of characters in the four English original films of *Little Women* and their Italian dubbed versions. Vignozzi paves the way for further exploration by reflecting on the efficacy of corpus linguistics in assessing the translation of multimedia texts. Following an overview of the big-screen adaptations of *Little Women*, he proceeds to detail the development of the corpus using Sketch Engine. Further, the analysis of the March sisters' speech was conducted in two stages, with a focus on the implicit textual cues identified by Culpeper's characterization model. Initially, a quantitative character-based analysis was performed by extracting keywords whose results were then subjected to a qualitative analysis. Besides, the concordance lines were examined to evaluate the translation of the dialogues.

In the last chapter of this book, Alessandra Rizzo investigates the linguistic features of dialogues and subtitles of TV crime dramas and their translations. She begins the chapter by situating this study within DH. A parallel corpus made from three episodes of three different TV crime dramas set in different geographical locations was compiled for the analysis. She then undertook a two-level analysis of the linguistic features of orality: one centred on examining language choices drawing on the theoretical framework of Halliday's Systemic Functional Linguistics (SFL) and the other on the translation strategies of linguistic features. Rizzo finally concludes with an interpretation of the outcomes and provides concluding remarks on the constraints and prospects of this study.

3 Critical evaluation

This book presents corpus linguistics and corpus-based translation studies as being within the purview of DH. It provides not only theoretical reflections on the burgeoning fields of research nested in DH, corpus linguistics, and translation studies with computers as pivotal components but also concrete case studies covering a wide range of research topics. Answering the call for the humanities to engage with digitalization, this seminal work touches the nerves of those seeking

to operate in this interdisciplinary or even transdisciplinary realm by paving the way for further discussion.

The biggest merit of the book is that it brings to the fore the nexus between corpus and DH and promises to consolidate the area. Although "corpus" and "corpora" are widely used in the literature on DH, it is surprising that corpus linguistics has been slow to embrace DH. A simple search with the keywords "digital humanities" in the SSCI & AHCI journals *International Journal of Corpus Linguistics* and *Corpus Linguistics and Linguistic Theory* indexed in CNKI Scholar academic database returns no results. It was only in the third decade of the 21st century that the relevance of DH to corpus linguistics began to be recognized. As a matter of fact, *The Routledge Handbook of Corpus Linguistics* (2020) dedicates a new chapter to corpora and DH, while in the 2010 edition, no instance of "digital humanities" is found.

In fact, this trend is also true of corpus-based translation studies. Tanasescu (2021) astutely highlights that it was only as recently as 2018–2019 that DH-inflected research started to gain more and more ground (in Translation Studies). Her observation coincides with the situation in China. Hu (2018) wrote an introductory article titled "Progress and Prospects of Translation Studies from the Perspective of Digital Humanities". In the same year, the Research Center for Digital Humanities led by Hongwu Qin, another distinguished scholar in corpus-based translation studies in China, was founded at Qufu Normal University. This edited volume resonates with the academic circle and will undoubtedly promote related research.

Another major merit is that it provides some case studies against which future research can be benchmarked. First, the wide-ranging case studies indicate what can be counted as DH. For example, Chapter 8 demonstrates the use of learning analytics in interpreter training, suggesting that the quantitative analysis of students' learning data is also a relevant component of DH. As online learning is becoming one of the most significant trends in educational settings (Mei et al., 2022), viewing learning analytics as being part of DH will shed light on the analysis of students' online learning data. Second, it provides procedural guidelines for this line of research, which usually comprise corpus design, data collection, data processing, concordance, and the interpretation of results with relevant theories, among others. Further, issues encountered during these processes and their corresponding solutions will be of reference value. For example, in Chapter 9, a mixed method of quantitative analysis and qualitative analysis is adopted to compensate for the limitations of each method. Third, it brings together many different areas of research such as legilinguistics, e-health, and films, allowing cross-referencing between these areas to adapt tools, methods, theories, and topics to their specific requirements. In summary, since the article by Jensen (2014) was published, other articles have touched upon the relationship between corpus linguistics and DH;

however, books dedicated to this topic with specific case studies have been few and far between. From this perspective, this book is a pioneering scholarly work that promises to inspire the adoption of DH.

Despite the aforementioned merits, this book is not without certain flaws. First, the title is slightly confusing. It seems to suggest that translation tools are included, but in fact, by tools, it means corpus tools for translation studies rather than computer-aided translation tools. Second, the relevance of DH to each chapter should be articulated explicitly. While some chapters provide discussions on the significance of DH, there are several chapters whose connections with DH are not explicitly stated, resulting in a certain degree of disjointedness between respective chapters and the volume as a whole; this is especially true of Part 1. Third, the tools and methods used in this book are still very limited and mostly confined to traditional corpus linguistics, as opposed to kaleidoscopic data processing and analysis tools and methods in DH. In the future, importance should be placed on adapting advanced tools and methods from DH and computational linguistics, among others, to meet the needs of this interdisciplinary field of research. Fourth, the connection between case studies and DH needs to be deepened. To be specific, DH has evolved into a field of research with domain-specific discourse comprising research topics and terms, among others. Future research carried out with DH in mind should incorporate DH discourse for cross-fertilization. Nevertheless, this book combines the corpus linguistics approach and the DH approach with humanities research, making it a ground-breaking seminal work for scholars of humanities in the digital age.

References

Hu, K. (2018). 数字人文视域下翻译研究的进展与前景 [Progress and prospects of translation studies from the perspective of digital humanities]. *Chinese Translators Journal*, *39*(6), 24–26.

Jensen, K. E. (2014). Linguistics and the digital humanities: (Computational) corpus linguistics. *Journal of Media and Communication Research*, *30*(57), 115–134.

Mei, F., Lu, Y., & Ma, Q. (2022). Online language education courses: A Chinese case from an ecological perspective. *Journal of China Computer-Assisted Language Learning*, *2*(2), 228–256.

Tanasescu, R. (2021). Complexity and the place of translation in digital humanities: Post-disciplinary communities of practice in the translation studies network. In K. Marais, & R. Meylaerts (Eds), *Exploring the implications of complexity thinking for translation studies* (pp. 30–72). Routledge.

Zheng, C., Yu, M., Guo, Z., Liu, H., Gao, M., & Chai, C. (2022). Review of the application of virtual reality in language education from 2010 to 2020. *Journal of China Computer-Assisted Language Learning*, *2*(2), 299–335.

Bionote

Fulu Liang
College of Foreign Languages, Nankai University, Tianjin, China
liang_fulu@126.com
https://orcid.org/0009-0005-1522-0342

Fulu Liang is a PhD candidate in Translation Studies at College of Foreign Languages, Nankai University. With a strong background in in-house translation spanning five years, he has gained extensive translation experience in industries such as metallurgy, automobiles, and wind power. His research interests are diverse, with a particular focus on the established field of technical translation and cutting-edge topics such as computational translation studies, digital translation studies, and language technology.