


# Book Review:

## Extending the Scope of Corpus-Based Translation Studies

Xiaowan Qin, Guangdong University of Foreign Studies, China\*

Keming Peng, Guangdong University of Foreign Studies, China

 <https://orcid.org/0000-0001-8975-3161>

### ABSTRACT

The field of corpus-based translation studies (CBTS) relies on corpus-linguistic methods and tools to analyze electronic corpora of authentic translations. Since its emergence in the early 1990s under the impetus of Mona Baker, CBTS has grown to become a recognized area of research in translation studies. By offering an overview of the field and presenting a variety of fresh perspectives provided by leading experts, The book *Extending the Scope of Corpus-Based Translation Studies* extends the scope of CBTS in multiple ways, and sheds light on the future of the translation industry.

### KEYWORDS

Corpus-Based Empirical Studies, Corpus-Based Translation Studies, Translator Training

The field of corpus-based translation studies (CBTS) relies on corpus-linguistic methods and tools to analyze electronic corpora of authentic translations. Since its emergence in the early 1990s under the impetus of Mona Baker (1993), CBTS has grown to become a recognized area of research in translation studies. Since 1995, Mona Baker and her team have started to build the first translational English corpus (TEC), and have conducted a series of corpus-based translation studies, including translation universals (Baker 1995, Laviosa 1998, Olohan and Baker 2000), translator's style (Baker 1999) and translation norms (Kenny 1998), etc. After the progress made in the past decades, it emerges clearly that CBTS now makes widespread use of theoretical insights and research methods borrowed from neighboring disciplines, such as translation process research, linguistic theory, contrastive linguistics, variational linguistics, contact linguistics, second-language acquisition and psycholinguistics (Kruger and Van Rooy 2016, Halverson 2017, Kruger and De Sutter 2018, De Sutter and Lefer 2020). Recently, some methodological advances have been evidenced by the meticulous description of the corpora, reflecting a huge leap in methodological rigor in CBTS (De Sutter et al. 2012), which in turn makes it possible to obtain much more solid insights into translation and other forms of bilingualism-influenced language varieties. By offering an overview of the field and presenting a variety of fresh perspectives provided by leading experts, *Extending the Scope of Corpus-Based Translation Studies* extends the scope of CBTS in multiple ways, and sheds light on the future of the translation industry.

DOI: 10.4018/IJTIAL.317105

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

The book, compiled by 16 authors in the field of corpus-based translations studies (CBTS), is a timely overview of the field today. The 9 chapters are organized into four parts, respectively focusing on “Corpus-Based Translation Studies: Current Challenges and Future Perspectives”, “Recent Methodological and Theoretical Developments in CBTS”, “Corpus-Based Empirical Studies” and “Corpus Use in Translator Training”.

The first part (Chapters 1 and 2) mainly provides a quantitative and qualitative overview of the current state of corpus-based research, and identifies dominant trends and potential weaknesses. Chapter 1 makes a survey based on relevant English scientific articles published in twelve scientific journals between 2012 and 2019, revealing that corpus approaches account for 11 per cent of the articles. Besides, a detailed investigation of the empirical studies, which are in the majority compared to methodology- and theory-oriented and applied studies, uncovers valuable information on the dominant research foci, corpus types and sizes, languages, modality and registers, as well as corpus techniques and statistical testing. Among other findings, the survey highlights a clear dominance of parallel corpora over monolingual comparable corpora, a strong focus on lexis and terminology, and a continued interest in translation features. It also shows a clear underuse of advanced corpus-linguistic and statistical techniques, a weakness that should be addressed in future research. Chapter 2 takes stock of the progress made in CBTS in the last three decades and suggests ways of expanding the reach of the field in future years. The author first discusses translation universals and translation directionality, and then discusses a key technological development for CBTS, namely the abundance of authentic digital translation data for many language pairs. In his view, it is now up to CBTS researchers to tame the digital data and devise new methods to tap into these potential corpus resources. As also shown in this chapter, CBTS can further extend its reach by examining other forms of translation than the ones traditionally included in parallel corpora, such as collaborative translation, translation crowdsourcing and video game localization.

The second part (Chapters 3 and 4) focus on recent methodological and theoretical trends in CBTS. Chapter 3 firstly discusses the rationale of the constrained-language framework, defines its key constructs and sets out its core theoretical principles, and then illustrates the variationist and multivariate method in a corpus study on the omission of the complementizer that in three varieties of English (English translated from Afrikaans, British English and South African English). While the varieties investigated are similar with respect to three constraint dimensions (register, proficiency and task expertise), they differ in the constraints of language activation (bilingual vs monolingual) and text production (dependent vs independent). The results of the random forest analysis and conditional inference tree modelling show that there are subtle, but significant, differences across varieties. While similar conventionality- and complexity-related factors are found to condition that-omission in the three varieties, corpus data show that translators opt for explicit that more often than writers in some specific registers and contexts. Chapter 4 deals with a key methodological development in corpus-based translation studies, namely the use of multiple methods. After an insightful discussion of how corpus, eye-tracking and keystroke logging data can be combined, the mixed-methods approach to translation is illustrated with a case study on subject identifiability in English and German. Taking as a starting point cross-linguistic differences in word order between English and German, the authors examine the role of subject identifiability on the positional shifts observed in translation. Their integrated analysis of corpus data, experimental product data and behavioral data confirms their initial hypothesis, namely that English non-identifiable subjects are more prone to translation shifts than identifiable subjects and that their translation is more effortful, possibly because they trigger a wider range of translation options from which to choose.

The third part (Chapters 5 to 7) features three corpus-based empirical studies, which use cutting-edge data extraction techniques, annotation systems and quantitative methods to provide innovative insights into translational products. Chapter 5 reports on a corpus-driven study within the constrained-language framework, which aims to assess the degree of commonality of constrained varieties that involve bilingual language activation. The study further innovates in controlling for

both constraining language and register, thus making it possible to disentangle cross-linguistic from register-dependent differences. Based on a method involving random forest-based keyness analysis and multidimensional analysis, the study brings out a complex interaction of constrainedness and register effects. Features shared by the constrained varieties include a tendency to rely on post-nominal modification and common nouns with determiners. Register variation proves to have a significant impact across varieties, although constrained varieties appear to be less register-sensitive than unconstrained ones. Chapter 6 deals with grammatical metaphors in translation. In systemic functional linguistics (SFL), a grammatical metaphor refers to a mismatch between meanings and their lexico-grammatical realizations. The authors identify the causes of translation shifts from grammatically metaphorical *of*-constructions in English to more explicit renderings in German. To do so, a range of source-language variables are examined, such as the semantic category and the context and co-text of the nominal *of*-constructions in two registers. Relying on a richly annotated corpus dataset and regression modelling, the authors show that the variables involved in the analysis do not influence translation shifts. This suggests that target language variables need to be considered to gain deeper insights into the phenomenon of de-metaphorization in translation. Chapter 7 examines the two phenomena of normalization and shining-through, that is, adherence to target-language versus source-language conventions, in both professional and student translations. Starting from a large set of lexico-grammatical features derived from variational linguistics and SFL, the author uses state-of-the-art automatic text classification techniques to trace normalization and shining-through in a multi-register English-to-German parallel corpus. Language conventions in English and German are modelled on the basis of comparable texts originally authored in the two languages. The results indicate that normalization and shining-through patterns do not vary across levels of translation experience, which goes against the author's initial hypothesis that novice translators should display lower register sensitivity than professionals. In future research, lexical features will need to be added to the current lexico-grammatical feature set to better characterize variation at the level of translation experience.

The fourth part (Chapters 8 and 9) provides two reports of classroom experiments of corpus-based translator training. The experiment in Chapter 8 aims to assess the relative impact of the use of a domain-specific monolingual target language corpus (MOC) versus the general bilingual concordancer *Linguee*. The author hypothesized that MOC use would result in a lower number of acceptability errors and a higher number of adequacy errors and that reverse results would be obtained for translations with *Linguee*. While the results for adequacy errors were confirmed, those related to acceptability were disconfirmed, the MOC condition resulting in a higher acceptability error rate, a finding that calls into question the positive impact of monolingual corpus use reported in some earlier studies. An analysis of error subtypes yields a wealth of interesting findings, such as the lower proportion of lexical errors in the MOC condition, showing that the debate on the use of monolingual versus bilingual resources is far from settled. The experiment in chapter 9 focuses on how students deal with the translation of complex noun phrases (CNPs) in specialized texts. The data consists of English-to-French translations produced by French-speaking translation students, in which all errors affecting CNPs have been annotated with the help of a fine-grained typology. The study aims to compare students' output under two task conditions: with the sole help of bilingual dictionaries and termbases versus a much wider range of resources, in particular two specialized English and French comparable corpora that students had learned to compile and query. A detailed analysis of the errors showed no difference in the number of errors produced with and without corpus, thereby casting further doubt on the all-round benefit derived from corpus use and pointing to the need for further research.

The merits of this book are manifold, and this review considers three of them worth emphasizing. First and foremost, this book showcases some of the latest trends in corpus-based translation studies. To begin with, an analysis of the overall corpus orientation of the studies into three main categories – empirical, methodological-theoretical and applied – showed that empirical studies accounted for two-thirds of the studies. In view of the descriptive, product-oriented slant of corpus linguistics, this

is not particularly surprising. What came as something of a surprise, however, and can be seen to testify to the growing maturity of the field, is that one-third of the studies went beyond description to tackle methodological and theoretical aspects and concrete applications. Second, a detailed scanning of the linguistic focus of each empirical study showed that the dominant category was that of lexis and terminology, followed by grammar, discourse and pragmatics, together with a mixed category comprising more than one linguistic domain. Semantic, speech-related and morphological features turned out to be less popular. Translation features (in particular, explicitation) proved to remain a popular subject of investigation, in line with Baker's research agenda. Third, the analysis of the corpus designs of the empirical studies showed that parallel corpora are used twice as frequently as monolingual comparable corpora, contrary to Baker's call to move away from ST-TT comparisons. Corpora used in the field were found to represent a wide range of written and spoken registers, with a clear overrepresentation of English (either as a source or as a target language). Fourth, methodology- and theory-oriented studies proved to be quite diverse, ranging from descriptions of new corpora, literature reviews and calls for the use of more advanced quantitative methods to the application of particular theoretical constructs or models, fostering cross-fertilization with neighboring disciplines. Finally, applied studies appeared to be mostly geared towards corpus use in translator and interpreter training, while other applied areas, such as corpus use in professional practice or translation quality assessment, were found to be rarely explored.

Secondly, the volume highlights the emerging interdisciplinary bridges between CBTS and other areas in linguistics and demonstrates the applications of these theories and methods to translator training. For example, it covers the latest theoretical developments, such as the constrained communication framework (chapter 3), with a strong focus on methodology, particularly mixed-method approaches, multivariate research designs and translation error annotation. Each of the studies discussed in chapters 3 and 4 followed the same general pattern combining corpus analysis with the integrated analysis of two behavioral methods. Moreover, each study had to cope with issues arising from the lack of an established methodology in empirical translation studies. Such issues include questions about the overall procedure, the timing and amount of detail of the corpus analysis, the generation and level of control of stimulus material, and so on. One lesson learnt from the original study discussed in chapter 4 is the added value of taking full advantage of each method. Methodological innovation is crucial in the process of establishing a discipline that previously did not draw on empirical research methods. For ascertaining empirical findings and deriving robust explanations, it is important to draw on a set of well-described and well-tested methodological procedures, which increase the explanatory power of empirical claims because they have withstood repeated rigorous tests. This is possible only with replicable designs. The discussions in Part II can contribute to the optimization of multi-strategy research and account for the influence of linguistic, cognitive, sociocultural and workflow-related factors on the final text consumed in a lingua-culture.

Another merit of the book, among many others, is that this volume examines lesser studied forms of translation, such as student translation and post-editing, and explores under-researched semantic and syntactic aspects of translated language, taking key variables into consideration, such as source language and register. In addition to showcasing recent methodological and theoretical developments in CBTS, this volume also reports on concrete classroom experiments on the use of corpora in translator training. Interestingly, the corpus component is twofold here, as corpus-based approaches to translation quality are adopted to assess the impact of corpus use by trainee translators. Even though electronic corpora are now widely used in translation curricula worldwide, empirical translation studies reporting on corpus-oriented teaching practices are rather rare. This is especially striking in recent edited volumes, where corpus applications, if discussed at all, are often limited to terminology and bilingual lexicography (Xiao 2010, Kruger et al. 2011, De Sutter et al. 2017).

The book is, of course, like many others, not perfect. Although the volume provides a comprehensive picture, corpus-based *Interpreting Studies* is still under-represented. This is due to the fact that corpus-based studies are currently less widespread in interpreting studies than in

translation studies, as indicated by the small number of interpreting studies presented at the UCCTS 2018 conference (10 per cent). Second, the picture drawn by the survey (Part I) is only partial as it is limited to journal articles written in English and therefore leaves out many relevant publications written in other languages and published in other formats. However, these issues do not compromise the value of this volume in extending the scope of corpus-based translation studies. Corpus-based translation and *Interpreting Studies* is still a relatively young research field. It is, therefore, only natural that some aspects of it have not yet attained full maturity. However, the fact that activity is thriving on all fronts – empirical, theoretical, methodological and applied – is a strong sign that the field will continue to progress unabatedly in the future. Thanks to its combined theoretical, methodological and applied perspective and innovative approaches, *Extending the Scope of Corpus-Based Translation Studies* will appeal to both seasoned specialists and newcomers to the field.

## DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

## FUNDING

This work is supported by the Fund of Key Research Projects of Humanities and Social Sciences in Colleges and Universities of Guangdong Province [2018WZDXM010] and the Fund of Graduate Research Innovation program of Guangdong University of Foreign Studies [22GWCXXM-108].

*Extending the Scope of Corpus-Based Translation Studies*, edited by by Sylviane Granger, Marie-Aude Lefer and Jeremy Munday, London, Bloomsbury, 2022, 288 pp., \$85.5 (hardback), ISBN: 9781350143258.

## REFERENCES

- Baker, M. (1993). Corpus Linguistics and Translation Studies. Implications and Applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233–250). Benjamins. doi:10.1075/z.64.15bak
- Baker, M. (1995). Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target*, 7(2), 223–243. doi:10.1075/target.7.2.03bak
- Baker, M. (1999). The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators. *International Journal of Corpus Linguistics*, 4(2), 281–298. doi:10.1075/ijcl.4.2.05bak
- De Sutter, G., Goethals, P., Leuschner, T., & Vandepitte, S. (2012). Towards Methodologically More Rigorous Corpus-Based Translation Studies. *Across Languages and Cultures*, 13(2), 137–143. doi:10.1556/Acr.13.2012.2.1
- De Sutter, G., & Lefer, M.-A. (2020). On the Need for a New Research Agenda for Corpus-based Translation Studies: A Multi-methodological, Multifactorial and Interdisciplinary Approach. *Perspectives*, 28(1), 1–23. doi:10.1080/0907676X.2019.1611891
- De Sutter, G., Lefer, M.-A., & Delaere, I. (Eds.). (2017). *Empirical Translation Studies: New Methodological and Theoretical Traditions*. De Gruyter Mouton. doi:10.1515/9783110459586
- Halverson, S. L. (2017). Gravitational Pull in Translation. Testing a Revised Model. In G. De Sutter, M.-A. Lefer, & I. Delaere (Eds.), *Empirical Translation Studies: New Methodological and Theoretical Traditions* (pp. 9–45). De Gruyter Mouton. doi:10.1515/9783110459586-002
- Kenny, D. (1998). Creatures of Habit? What Translators Usually Do with Words. *Meta*, 43(4), 515–523. doi:10.7202/003302ar
- Kruger, A., Wallmach, K., & Munday, J. (Eds.). (2011). *Corpus-Based Translation Studies: Research and Applications*. Bloomsbury.
- Kruger, H., & De Sutter, G. (2018). Alternations in Contact and Non-Contact Varieties: Reconceptualising That-Omission in Translated and Non-Translated English Using the MuPDAR Approach. *Translation, Cognition & Behavior*, 1(2), 251–290. doi:10.1075/tcb.00011.kru
- Kruger, H., & van Rooy, B. (2016). Constrained Language: A Multidimensional Analysis of Translated English and a Non-Native Indigenised Variety of English. *English World-Wide*, 37(1), 26–57. doi:10.1075/eww.37.1.02kru
- Laviosa, S. (1997). How Comparable Can ‘Comparable Corpora’ Be. *Target*, 9(2), 289–319. doi:10.1075/target.9.2.05lav
- Olohan, M., & Baker, M. (2000). Reporting That in Translated English: Evidence for Subliminal Processes of Explication? *Across Languages and Cultures*, 1(2), 141–158. doi:10.1556/Acr.1.2000.2.1
- Xiao, R. (Ed.). (2010). *Using Corpora in Contrastive and Translation Studies*. Cambridge Scholars Publishing.

*Keming Peng is a PhD at the School of Interpreting and Translation Studies, Guangdong University of Foreign Studies and Research Assistant at Guangdong Institute for International Strategies. He used to work in the United Nations as a translator and is now a National Consultant (Translator) of United Nations Development Programme (UNDP).*