East Asian perspectives on corpus-based interpreting studies

Andrew K. F. Cheung, Kanglong Liu, and Riccardo Moratto

Corpus-based translation studies (CBTS) have firmly established their position in academic discourse as a mature discipline that exerts a profound influence on the research landscape of translation studies since Baker (1993) put forward the idea of applying corpus linguistics in translation studies. Employing corpus-linguistic approaches and computer-aided methods to analyze extensive collections of translated texts, CBTS has facilitated a nuanced understanding of a wide range of translation phenomena (e.g., Kruger and van Rooy 2016; Liu and Afzaal 2021; Xiao and Hu 2015). Through systematic exploration of language usage patterns in translation, CBTS has significantly expanded our comprehension of the intricate dynamics inherent in the translation process.

Despite the prominence of CBTS, the application of similar methods in interpreting studies has faced a slower adoption. In 1998, a seminal article entitled *Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies* was published. Shlesinger (1998) pioneered the potential use of a corpus-based approach to examining interpreting and discussed its benefits and challenges. While interpreting and translation share common ground as language mediation activities, interpreting possesses distinct characteristics that warrant separate methodological considerations.

The emerging discipline of corpus-based interpreting studies (CBIS) has encountered its fair share of challenges. One major obstacle revolves around the demanding task of constructing spoken corpora. Unlike written texts, spoken language is transient and context-dependent, posing unique complexities in capturing, transcribing, and annotating oral data (cf. Shlesinger 1998; Love 2020). This process demands substantial resources, specialized skills, and sophisticated technology, thereby presenting significant hurdles for researchers aiming to develop interpretable corpora. According to Burnard (2002, 58), "A rough guess suggests that the cost of collecting and transcribing in electronic form one million words of naturally occurring speech is at least ten times higher than the cost of adding another million words of newspaper text." Besides, the predicament lies in identifying and utilizing the most effective indicators for analysis amidst the vast array of potential features within spoken corpora. Distinguishing the most salient and informative factors among the

DOI: 10.4324/9781003377931-1

many variables that characterize spoken language is a formidable endeavor. Also, the specific nature of interpreting events, including the intricate interplay among interpreters, speakers, and audiences, further complicates this task, necessitating the development of innovative analytical approaches. It is noteworthy that there has been a relative lack of interest from professional interpreters in conducting research, particularly when employing corpus-linguistic methods. Given that interpreters excel in practice and performance, their potential detachment from the academic aspects of interpreting is not unexpected, as their expertise lies more in practical experience and knowledge than in empirical analysis or theoretical exploration. However, positive change is underway. CBIS is undergoing a significant transformation and has become an important research direction (Pöchhacker 2022, 42). Researchers in the field of interpreting studies are increasingly recognizing the significant benefits of using corpora to unveil concealed aspects of interpreting that are often overlooked by traditional qualitative or experimental methods (Cheung 2017; Liu, Cheung, and Liu 2023). Unlike these conventional approaches, which typically gather limited or decontextualized data (Wang and Tang 2020), incorporating corpora enables a more holistic and contextual comprehension of interpreting.

The growth and maturation of CBTS have paved the way for CBIS, offering established methodologies, research designs, and analytical methods that can be adapted and applied to the realm of interpreting studies. However, the current research landscape in interpreting studies exhibits a noticeable limitation in its scope, as it primarily focuses on European languages and contexts (Setton 2011). This narrow focus is particularly evident when considering the data sources predominantly retrieved from institutions such as the European Parliament, the European Commission, and the European Council (Russo, Bendazzoli, and Defrancq 2018). This bias has resulted in a scarcity of research encompassing East Asian languages such as Chinese, Japanese, and Korean. Addressing this omission holds significant importance for the ongoing development of CBIS for two compelling reasons. First, East Asian languages possess typological distinctions from their European counterparts, suggesting that their study can provide valuable methodological insights and unveil phenomena that may remain elusive in European languages. Second, the East Asian context diverges from its Western counterparts in terms of cultural, social, and institutional norms, which in turn may influence interpreting processes and outcomes. Thus, exploring interpreting phenomena within the East Asian context holds the potential to shed light on rich explanatory dimensions that differentiate it from Western contexts.

Recognizing these pressing needs, we have compiled this edited book to bridge the gap in the field and shed insights on CBIS in the East Asian context. Our primary aim is to push the boundaries of CBIS and make meaningful contributions to its burgeoning scholarship, with a particular focus on the dynamic East Asian sphere. This comprehensive volume encompasses a diverse range of topics, embodying the rich diversity and dynamism of the interpreting

field. It delves into multiple facets of interpreting, exploring captivating areas such as medical interpreting (Li and Cheung), court interpreting (Wong; Xue), political interpreting in the United Nations Security Council (Wen and Al-Bekai), and interpreter training (Choi; Lee and Liao), offering a kaleidoscope of perspectives. The languages examined within these contributions include Chinese-English/English-Chinese (Gu and Wang; Ma; Qi and Yang; Zhang), English-Japanese (Yamada, Matsushita, and Ishizuka), Chinese-Portuguese (Vital and Han), and Cantonese-English (Cheung Pease and Pease; Xu and Liu), presenting a unique cross-linguistic lens through which to explore the intricacies of interpreting processes. The wide spectrum of linguistic diversity showcased in these chapters underscores the robustness and versatility of corpus-based methods, demonstrating their applicability across a diverse array of language pairs. Our contributors hail from across the East Asian region, encompassing the Chinese mainland, Japan, Korea, Hong Kong, Taiwan, and Macao. This regional representation serves as a testament to the vibrancy and diversity within the realm of interpreting studies in East Asia. Their contributions offer multifaceted cultural and contextual perspectives, thereby broadening the horizons and enriching the current understanding of CBIS.

Each chapter within this volume, while unique in its focus and methodology, converges toward a shared objective: to deepen our understanding of interpreting phenomena through the application of corpus-based methods. This book extends beyond being a mere compilation of studies; it serves as an open invitation to reflect upon the practice and study of interpreting in East Asia, to challenge existing paradigms, and to explore new avenues of research. In addition to the empirical insights offered by each chapter, this book underlines the pressing need for continuous innovation in methodological approaches within CBIS. It emphasizes the urgency of developing strategies for constructing, managing, and analyzing spoken corpora, identifying pivotal indicators within corpus data, and enhancing the involvement of professional interpreters in research endeavors. Our hope is that this book will serve as a catalyst for the exploration of new contexts, the integration of diverse research traditions, and the creation of synergies across various areas of inquiry. Hopefully, CBIS can finally reach "a stage in its development as a discipline when it is ready for and needs the techniques and methodology of corpus linguistics in order to make a major leap" (Baker 1993, 248). We express our profound gratitude to the contributors for their rigorous work and unwavering passion for corpus-based interpreting studies. Their commitment to exploring interpreting in the East Asian context has greatly contributed to the significance of this book for the advancement of the field.

This book comprises 15 chapters that delve into the application of corpora in studying various interconnected issues within interpreting. To facilitate reading, abstracts of the authors will be presented below, providing concise summaries of each chapter.

In Chapter 1, Hao Yin, Han Xu, and Kanglong Liu comprehensively review corpus-based interpreting studies (CIS) in China during the past decade

(2013–2022) and identify emerging research directions. Through a systematic examination of influential papers published during that period, a flourishing phase of advancement in CIS is observed. Specifically, this chapter reveals a substantial expansion in the research scope, primarily attributed to advancements in techniques for compiling, transcribing, and annotating interpreting corpora. The use of self-built corpora and the adoption of diverse research methodologies have been pivotal factors driving the significant expansion observed. In addition, the current state of CIS in China reveals notable trends, including a dominance of synchronic studies and an emphasis on the interpreting process and discourse-level aspects. Quantitative analysis is increasingly employed, and the expansion in the number and scope of interpreting corpora indicates a more diverse landscape. However, limitations exist, such as a lack of diversity in corpus sources, immaturity of transcription and annotation techniques, as well as an imbalance in the availability of different types of corpora. Addressing these limitations will promote a comprehensive understanding of CIS in China.

In Chapter 2, Jennifer LF Cheung Pease and Adam Pease study the speech act of clarification in the context of courtroom interpreting. Clarification, although not the focal point of dialogue, plays a vital role as an aside. To address the research gap in this area, the authors conduct a quantitative analysis of clarification based on a discourse analysis framework of clarification speech acts, using formal ontology at the communication level. They categorize the pattern of clarifications in an interpreter-mediated courtroom in Hong Kong and count their frequency. The statistical results validate some previous observations about clarification and uncover potential new explanations for such instances of clarification. The authors also examine dialogue repair resulting from clarification. In this chapter, the authors demonstrate how the framework of clarification analysis can be served as an effective tool for studying the difficulties faced by interpreters of Asian languages and evaluating their performance.

In Chapter 3, Zi-ying Lee and Min-Hsiu Liao present a corpus-based study of trainee interpreters' reflective journals. The study aims to systematically explore thematic issues related to the concerns and difficulties encountered by interpreting students, as well as their perception of instructors' and peers' roles during their learning process. Drawing inspiration from learner-centered approaches in education studies, reflective journals have become not only a common practice in interpreting classes but also an emerging research field in interpreting studies. This study compiles a corpus of reflective journals from undergraduate and postgraduate students taking interpreting courses at a British university. The analysis focuses on identifying the general ideational themes derived from the word frequency list in the corpus and examining the agents involved in the students' reflective narratives using a specific lexical search. The findings are then triangulated with the interpreter assessment criteria and the curriculum design of the students' interpreting classes to provide a better contextual understanding. The authors emphasize that while thematic

and content analyses enable the identification of important issues, a corpus-based approach can complement these studies by revealing consistent patterns of specific linguistic features across a large dataset.

In Chapter 4, Yifan Zhang examines the frequency and pragmatic function of five types of fuzzy language in the English output of a Chinese-English (C-E) interpretation of the Chinese premier's press conferences, using the C-E interpreting corpus. The study results show three critical findings. First, the frequency of fuzzy language is reduced in the output of the C-E press conference interpretation. Among different types of fuzzy language, fuzzy words have the highest density. Second, five types of fuzzy languages generate seven kinds of pragmatic functions, which mainly aim to improve the quality of the interpreted output and the efficiency of communication. Third, besides the pragmatic function of face-saving, the remaining functions can be achieved by one or two types of fuzzy language. Overall, this chapter contributes to the understanding of how different types of fuzzy language serve various pragmatic functions, particularly in press conference settings.

In Chapter 5, Xingcheng Ma investigates the use of strategy for coping with word order asymmetry by professional interpreters during simultaneous interpreting with text (SIT) and without text (SI). With respect to whether the original word orders were changed in the target production, two major strategies, that is, chunking and reordering, were identified for rendering the asymmetric structure. A corpus-based descriptive approach was adopted based on the source input and target output of English-to-Chinese simultaneous interpreting during sessions of the United Nations General Assembly. A binary logistic regression analysis was applied to explore the interpreters' strategy choice between chunking and reordering for relative clauses (RC) under SIT and SI. This analysis allows for a holistic examination of both the mode-specific effect and linguistic variables indicative of the cognitive constraints during interpreting. The results demonstrate a significantly greater occurrence of reordering in SIT. Strategy preference was also impacted by syntactic parameters of RCs including RC length, length of the head word and RC type. This study confirms the relationship between interpreting mode and the cognitive processing of word order asymmetry. Furthermore, it reveals a notable effect of syntactic features of relative clause on interpreters' strategic decision-making.

In Chapter 6, Taoyun Qi and Cheng-shu Yang explore the universal features in professional interpreters' English-Chinese simultaneous interpreting (SI) from the perspective of frequency of pause, one of the most important prosodic phenomena. The authors collected 12 pairs of SI audio files by 12 different professional interpreters, with each pair consisting of audio clips of both the source and interpreted speeches. A multimodal corpus was built using the software Praat in which pauses were marked with 0.250 seconds as the threshold duration value. The source text (ST) and target text (TT) were presented in a time-aligned manner with Praat's multilayered annotation function. The likelihood ratio test indicates that the pause frequency in TT is no lower than

that in ST, and that characteristic, as the subsequent qualitative multimodal analysis finds, is caused by the overwhelmingness of the interpreter's pause-increase-operations against the pause-decrease-operations. Seven patterns of pause-increase-operation and four patterns of pause-decrease-operation are extracted. The overwhelmingness of the pause-increase-operations reflects the remarkable command of SI expertise by professional interpreters.

In Chapter 7, Danni Li and Andrew K. F. Cheung discuss the common practice of hiring non-professional interpreters at medical conferences in mainland China. These interpreters, often health professionals with medical knowledge and language skills, are preferred over professional interpreters. This raises questions about the professionalism of simultaneous interpreting at medical conferences. The chapter examines the use of modal verbs by professional and non-professional interpreters at online medical conferences. The findings suggest ways to improve professionalism in this field.

In Chapter 8, Jiaqi Xue emphasizes the role of interpreter as an important topic for studies in court interpreting. In this chapter, Goffman's participation framework was utilized to investigate the role of interpreter based on a corpus of Hong Kong court interpreting. The obtained results show that in Hong Kong court interpreting, interpreters generally play the role as animator. To be more specific, in Cantonese to English rendition, interpreters mainly act as animator while author is mostly exhibited during the process of English to Cantonese rendition. In particular, there is a huge increase in the role of principal in English to Cantonese rendition in contrast with Cantonese to English rendition. In this study, a comprehensive and systematic description of the role of court interpreter in Hong Kong is provided, which may serve as a different research perspective in Hong Kong court interpreting.

In Chapter 9, Ting-hui Wen and Wassim Al-Bekai compare the differences between native and non-native English interpreting. The authors compile a corpus of United Nations Security Council interpreting, comprising two comparable subcorpora: 1) interpreting from French into English by native interpreters and 2) interpreting from Chinese into English by non-native interpreters. The results show that native interpreters had a relatively higher standardized type-token ratio than non-native interpreters, indicating higher lexical diversity. Moreover, native interpreters tended to use more nominalization, demonstrative devices, and pronominal devices than non-native interpreters, especially first-person pronouns. Finally, a wordlist created using WordSmith showed that fixed expressions such as "in that regard" and "with regard to" were used more frequently by native interpreters than by non-native English interpreters.

In Chapter 10, Cecilia Lok Yee Wong examines the types and frequencies of non-renditions in a 100-hour corpus of court interpreting records from Hong Kong courtrooms. Interpreters' non-renditions are utterances that do not have a source language which interpretations generate from (Wadensjö 1998). The assumed roles of the interpreters within the courtrooms are relative, however they embody immense power. Professional boundaries are overstepped

through interactions by the interpreters particularly toward lay participants. The distorted distribution of non-renditions, frequency to question, and direct lay participants communicated through non-renditions suggest court interpreters' manifestation of power in the courtrooms. Social regulations and social boundaries inform our actions and behaviors (Goffman 1959; French and Raven 1959). In a courtroom context, this study predominantly focuses on the contextual functions of non-renditions. The manipulation of power, control and influence embodied by the interpreter in a courtroom is evident.

In Chapter 11, Gu and Wang argue that interpreting studies have traditionally focused excessively on the ST, neglecting the importance of the interpreting product (TT). This is particularly evident in studies that emphasize interpreting quality and accuracy. The chapter asserts the significance of the interpreting product as a vital source of historical, socio-political, and diplomatic knowledge, regardless of its fidelity to the original. In today's interconnected world, the words spoken by interpreters in high-profile and high-stakes situations carry weight and cannot be undone. Therefore, the interpreting product holds greater international currency and global relevance. The chapter calls for research that goes beyond individual linguistic features and categories, emphasizing the need to analyze the propositional content and ideational elements conveyed in interpreting. It advocates for viewing the interpreting product as a corpus and a legitimate source of empirical data, offering interdisciplinary insights for researchers in various fields such as social and political sciences, history, foreign affairs, and diplomatic studies. The study itself focuses on China's self-representational discourse, specifically examining the collocation "China is," using a corpus of political interpreting data spanning a significant period.

In Chapter 12, Han Xu and Kanglong Liu explore the manifestation of simplification in interpreting based on a comprehensive corpus consisting of English speech simultaneously interpreted from Chinese (interpreted speech), English speech produced by native speakers (L1 speech), and English speech made by advanced non-native speakers (L2 speech). The level of simplification is quantified using four lexical complexity indicators: type-token ratio, lexical density, list head, and core vocabulary coverage. While not all the lexical indicators consistently confirm the occurrence of simplification in interpreting, there is a tendency for interpreters to rely on less varied lexical options and more high-frequency words, partially supporting the hypothesis of simplification. In addition, the findings reveal that simplification is not unique to interpreting, as an even more pronounced manifestation was observed in L2 speech. Considering the highly cognitively demanding nature of producing interpreted and L2 speech, the occurrence of simplification may result from the producers' efforts to alleviate cognitive load.

In Chapter 13, Ivo Vital and Lili Han argue that corpus linguistics has been revealing its methodological relevance in language-related studies. By examining corpus use in interpreting, this chapter adopts a thematic corpus from the Sketch Engine platform for the purpose of conference interpreting preparation,

to study its influence on the quality of consecutive interpreting, with a special focus on Chinese-Portuguese language combination. Basing on experimental protocols designed by previous studies (Díaz-Galaz 2012; Xu 2015, 2018), a pilot-study was designed and conducted, taking into account the specificities of the Chinese-Portuguese language combination in Macao's circumstances, to explore the synergetic influence of terminological preparation for interpreting. A group of MTI students from Macao participated in the experiment, and both test group and control group were created for data analysis. Preliminary results manifested that, despite the positive influence of corpus preparation on the quality of interpreting, its conspicuity was more noticeable when it was combined with other preparation methods. This conclusion goes beyond the experiment's scope, but contributes to an integrated conference interpreting preparation design for interpreting curriculum and training.

In Chapter 14, Masaru Yamada, Kayo Matsushita, and Hiroyuki Ishizuka delve into the establishment and analysis of an interpreting corpus employing remote simultaneous interpreting (RSI), with specific attention given to English-Japanese language pairs. Their research team engaged a group of 12 professional interpreters, whose task was to interpret a selected press conference. This process was not only useful for data gathering but also provided an opportunity to compare the conventional on-site interpreting with RSI interpretations. After the collection phase, the study proceeded with the transcription and bilingual alignment of data. To ensure alignment accuracy, especially between English and Japanese texts, a customized Vecalign algorithm was utilized. Additionally, the research critically examined the applicability of the Multidimensional Quality Metrics (MQM) for assessing interpreting performance. Notably, the findings indicate that MQM, while comprehensive, may not fully capture the intricacies of simultaneous interpreting. The study unearthed a high percentage of accuracy errors, primarily omissions and meaning errors. These findings underscore the need for the development of more tailored evaluation criteria designed specifically for interpreting performance.

In Chapter 15, Moonsun Choi argues that speeches or STs are important in interpreter training. However, there has been a relative dearth of information available on the characteristics of the materials used by interpreting teachers and students in training. This study attempts to shed light on the textual characteristics of English STs used in professional interpreter training programs in Korea and China. STs from the two countries were compiled into corpora. The study first examined the basic descriptive statistics of the corpora and compared them against large reference corpora to understand their general characteristics. Second, keyword analysis was performed to identify the main content of STs and gain insight into the topical focus of interpreter training in Korea and China. Lexical features of STs were explored from the perspective of the composition of high- and low-frequency vocabulary and lexical variation by means of text coverage and the type-token ratio. Finally, STs were examined and compared on multiple dimensions of text difficulty. The findings of this study contribute to an enhanced understanding of the

materials used for training interpreters in Korea and China and further help inform and advance interpreter training programs in other countries by providing a useful reference point for material selection and development.

References

Baker, Mona. 1993. "Corpus Linguistics and Translation Studies: Implications and Applications." In *Text and Technology: In Honour of John Sinclair*, edited by Mona Baker, Francis Gill, and Elena Tognini-Bonelli, 233–50. Philadelphia: John Benjamins.

Burnard, Lou. 2000. "Where did we Go Wrong? A Retrospective Look at the British National Corpus." In *Teaching and Learning by Doing Corpus Analysis*, edited by Bernhard Kettemann and Georg Marko, 51–70. Leiden: Brill.

Cheung, Andrew K. F. 2017. "Non-renditions in Court Interpreting: A Corpus-based Study." *Babel* 63(2): 174–99. https://doi.org/10.1075/babel.63.2.02che.

Díaz-Galaz, Stephanie. 2012. "La influencia del conocimiento previo en la interpretación simultánea de discursos especializados: Un estudio empírico." PhD Thesis, Universidad de Granada.

French, John and Bertram Raven. 1959. "The Bases of Social Power." In *Studies in Social Power*, edited by Dorwin Cartwright, 150–67. Ann Arbor, MI: Institute for Social Research.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday.

Kruger, Haidee and Bertus van Rooy. 2016. "Constrained Language: A Multidimensional Analysis of Translated English and a Non-native Indigenised Variety of English." *English World-Wide* 37(1): 26–57. https://doi.org/10.1075/eww.37.1.02kru.

Liu, Yi, Andrew K. F. Cheung, and Kanglong Liu. 2023. "Syntactic Complexity of Interpreted, L2 and L1 speech: A Constrained Language Perspective." *Lingua* 286: 103509. 10.1016/j.lingua.2023.103509.

Liu, Kanglong and Muhammad Afzaal. 2021. "Translator's Style through Lexical Bundles: A Corpus-driven Analysis of Two English Translations of Hongloumeng." *Frontiers in Psychology* 12: 633422. https://doi.org/10.3389/fpsyg.2021.633422.

Love, Robbie. 2020. *Overcoming Challenges in Corpus Construction: The Spoken British National Corpus 2014*. New York: Routledge.

Pöchhacker, Franz. 2022. *Introducing Interpreting Studies (3rd Ed.)*. Abingdon: Routledge.

Russo, Mariachiara, Claudio Bendazzoli, and Bart Defrancq. 2018. *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer.

Setton, Robin. 2011. "Corpus-Based Interpreting Studies (CIS): Overview and Prospects." In *Corpus-Based Translation Studies Research and Applications*, edited by Alet Kruger, Kim Wallmach, and Jeremy Munday, 33–75. London: Continuum.

Shlesinger, Miriam. 1998. "Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies." *Meta* 43(4): 486–93. https://doi.org/10.7202/004136ar.

Wadensjö, Cecilia. 1998. *Interpreting as Interaction*. London: Longman.

Wang, Binhua and Fang Tang. 2020. "Corpus-based Interpreting Studies in China: Overview and Prospects." In *Corpus-based Translation and Interpreting Studies in Chinese Contexts: Present and Future*, edited by Kaibao Hu and Kyung Hye Kim, 61–87. Cham: Palgrave Macmillan.

Xiao, Richard and Xianyao Hu. 2015. *Corpus-Based Studies of Translational Chinese in English-Chinese Translation*. Berlin: Springer.

Xu, Ran. 2015. "Terminology Preparation for Simultaneous Interpreters." PhD Thesis, University of Leeds.

Xu, Ran. 2018. "Corpus-based Terminological Preparation for Simultaneous Interpreting." *Interpreting* 20(1): 29–58.