Xinchao Guan and Xiaolei Lu (2022). Python-based Corpus Translation: Data Analysis and Theoretical Exploration. Shanghai Jiao Tong University Press. 223 pp.

"Python-based Corpus Translation - Data Analysis and Theoretical Exploration", authored by *Xinchao Guan* (Shanghai Jiaotong University) and *Xiaolei Lu (*Xiamen University*)*, published by Shanghai Jiaotong University Press, responds to the interdisciplinary integration in the context of artificial intelligence. In general, this book aims to provide theoretical insights and technological innovation in the direction of corpus translation based on Python programming language, which involves the corpus translation in both the academic field and the business industry.

As is known to all, machine translation in the field of corpus translation has access to most applications but the quality needs to be improved in certain scenarios. In addition, corpus translation has witnessed the lack of corpus retrieval and analysis tools, inconsistency of software coding format, insufficient accuracy of term extraction, and problems in the integration and application of technical tools. And that's where Python has a role to play. As an independent and powerful programming tool, Python can be and should be integrated into related research and practice of corpus translation with loads of advantages especially compared with other similar third-party tools. In fact, it is this very book carried with the academic research that narrows down the complicated functions of Python into the fields of language-corpus translation which make it possible, vivid and beneficial to most readers.

Although several python-and-translation related books that are relevant to corpora have been published recently around the world, the contribution of "Python-based Corpus Translation" is unique in at least three ways. First, it not only gives the steps of Python programming for free, but also shares the pedagogy of integrating translation technology into translation classes. Second, the book provides a wide variety of first-hand Chinese to English and English to Chinese translation models based on a large amount of translation teaching, translation theory research and students' translation practice in law, politics, literature and other fields, which are especially friendly to readers who are seeking for eye-opening ideas, cases and tools. Third, this book compares alternative approaches and describes the limitations of different methods of corpus translation.

The book comprises 223 pages structured into 9 chapters: corpus translation overview, multi-variable synergy, readability of translation, phrase features of translation equivalence, construction and application of translation knowledge base, language structure recognition and post translation editing, translation quality evaluation, and text information, mood comparison analysis and development and application of bilingual tools

The first part, i.e. Chapter 1, is a comprehensive overview and analysis of the academic research and industrial application of corpus translation. First of all, the concept of corpus translatology, multi-variable setting, and the basic concepts of corpus and translation in the industry are given, so that readers can get snapshots of the past and present of the development of this discipline. Even if they come across definitions such as corpus and translation, they can quickly understand the relationship between linguistic features and contextual variables.

Further chapters (2-8) present in detail different types of python-based models and variables that have proven helpful for corpus translation. It kicks off translation theory as the starting

point to explore the possibility of integration of translation theory and current technology, involving specific bilingual comparisons.

Chapter 2 introduces different variable settings focused on the translator's case studies. How to extract source texts and target texts, how to analyze words and phrases, how to express ideas and how to present interlingual variables are clearly illustrated here. In particular, a specific case of the English translation practice of the Copyright Law of the People's Republic of China is given, from which the multiple linear regression of lexical dimensions and the comparison of semantic transfer of bilingual texts are discussed, highlighting the profound foundation of statistical methods, and combining a certain amount of visual charts, equations and models. With the extraction results and calculation results are given, most readers can understand the multidimensional analysis method in the field of corpus translation at a glance. In general, this case adopts five test methods, which shows a rigorous scientific attitude of the author and sets an academic model for science studies of the translation research and development of digital humanities.

Chapter 3 is about the readability of Text/Translation. Hereby, Dale Hall Index formula and Flesh Kincaid Grade formula are presented accordingly. Following the case of "Copyright Law" in the previous chapter, we could get a better understanding of the applicability and readability of formulas into professional corpus translation, such as some formulas are more applicable to military and technical texts. And the comparison of sentence complexity of the translated text shows the relevant exploration path.

Chapter 4 mainly covers the concept of translation equivalence. This chapter is friendly for teachers and students of linguistics and translation studies. It discusses four types of translation equivalence, pointing out that Python has multiple paths to achieve translation equivalence in terms of passage, far beyond terminology, which also draws a more important conclusion: "The more stylized a text is, the more likely it is to be automated." In addition, a bilingual parallel corpus is created based on the text of the United Nations Convention, exploring the path of phrase equivalence in translation.

Chapter 5 focuses on translation knowledge database. It mainly introduces semantic knowledge pool, sharing knowledge-based translation, especially the practice of terminological expression in medical and financial fields (English to German), knowledge expression in the era of machine translation, such as polysemes “normal” in English and “Ze Ren” (meaning responsibility) in Chinese, which is interesting, detail-oriented and full of tension.

Chapter 6 serves as the middle section of the whole book, explaining language structure recognition and MTPE (i.e. machine translation and post editing). This chapter brings two major cases on the most welcomed topics- “medicine” and “oversea publicity” among researchers and citizens. What impressed me most could be the analysis of "Alzheimer's Disease", combined with "Scientific Translation”. The writers compared the machine translation of Baidu, Google, Sogou and other multiple translations to interpret the density structure of function words. The "News Discourse" section reflects the practical role of Python-based corpus translation in political publicity in the context of the current era of "Telling Good Stories of China", aiming to assist the academic circle and the industry to build national discourse from the technical level. In a much broader sense, these case studies present the unexpected side of Python-based Corpus Translation. It is of epoch-making significance to promote the social responsibility and responsibility of today's scholars beyond classroom teaching practical. It is this part that could amaze or inspire the readers to be ambassadors of

“Branding Wonderful Voices of China” as part of the nation’s strategy in the era of digital humanities.

In Chapter 7, the translation quality assessment recommends the most advanced automatic assessment methods at home and abroad. With the formula charts and statistics, it shows a key role played in the translation quality management process. This chapter refers to the important concept of computational linguistics, “Word Vector". On top of that, based on the translation of "Company Law" as the research text, it clarifies the complete programming, extracting and coding with analysis in a practical way. Owing to the vast amount of translation quality assessment techniques, this chapter has carried out a preliminary introduction and played a fine guiding role, serving as literature review.

In Chapter 8, the comparative analysis of text emotion is given, including bilingual emotion analysis under mixed mode and comparable relationships as the text with emotional bias is mined through technology. From the lexical level and grammar level, it reorganizes the exploration overview of translation relations and emotional analysis, so as to explain its social and cultural significance. What is worth mentioning, this chapter takes “2020 Annual Report” of the top-notch technology giant “Huawei Investment Holding Co., Ltd.” as an example to analyze the phenomenon of "addition" in translation, a technique that MTI students and teachers often attached attention to. The quantitative analysis of translation of the “Baowei Zhan” (meaning”war of defense" or “battle”) is also presented. Apart from this, based on a white paper called “the Chinese Action against the COVID-19 Epidemic” as the focus of the whole country, detailed theories are explained with texts and data. What is more unique is that the visual results show the concrete performance of DeepL- driven machine translation in the aspect of emotional translation transmission, attracting readers to give it a second thought.

The thought-provoking Chapter 9 is the third part and presents a variety of bilingual Python technology solutions, focusing on the development and application of bilingual tools, aiming to expand the integration of Python and corpus translation. For instance, it brings the ParaCone bilingual retrieval function, and compares Python+Excel mode with others. This chapter also introduces bilingual parallel alignment and parallel text format exchange, as well as innovative tools, such as parallel text information extraction methods and bilingual text cleaning, which leaves the readers panting for more.

In terms of creativity, this book is bold in putting things into practice and blazing new trails with three major characteristics. For a start, it lists a large number of Chinese and foreign literature reviews, autonomous programming techniques and the latest translation practice cases. In addition, the translation research knowledge and translation practice experience of the author is transformed via the application of technical programming. Thirdly, Python technology is more specific, highly practical, and easy to understand, which is more conducive to the promotion of corpus translation and research, and can serve as a guidance or reference book for all.

In a word, the book includes both theoretical description and practical application, emphasizing the feasibility of current theory and practice together. It is not only suitable for translation enthusiasts, but also for students and pedagogy guidance. In the age of big data, this book is a secret pathway to efficiently manage language assets and customize the translation knowledge database, even if it takes the Chinese and English parallel corpus as examples. Mastering the creation, processing and retrieval of corpus and terminology are the basic requirements for professional translators and professors.

My main suggestion for the next edition is that more emphasis may be put on case studies to involve a wider range of translators both home and abroad. I have much praise for this book as it deserves to reach a broader audience. As a paradigm of the localization of digital humanities across the whole world, this books is calling for a new discipline-"computational translation" that will come into being in the future as the author implies. It's the trend of the times that the artificial intelligence technology and humanities research will be merged into each other in the age of big data. This is not merely measurement research, but a hybrid research of qualitative and quantitative collaboration. Last but not the least, an entirely new era is beckoning us on.

Xiaolin Xie
Northeastern University, China
(Email: 37256685@qq.com)

Received: 1 September 2022; Accepted: 1 December 2022; Published: 16 May 2023
https://doi.org/10.58304/tc.20230206

Xiaolin Xie, a R.A. and T.A. of MTI at Northeastern University, a TESOL and TEFL scholar, reviewer and editor of some international journals and a CATTI lecturer. She's been focusing on research and consulting fields including an academic journey in the US. She also holds a Master Degree of Advertising Science at Communication University of China and a Bachelor Degree of Linguistics and Applied Linguistics at Dalian University of Foreign Languages. Her major research concerns translation, interpretation, corpus, linguistics, digital humanities as well as discourse, media, news and journalism, marketing, branding, communication and pedagogy.