The Construction of Parallel Corpus Based on the Seven Classical English Versions of *Tao Te Ching*

Xincun Huo[1, *], Fumin Zuo[2]

[1] *School of Humanities, Arts and Education, Shandong Xiehe University, Jinan, Shandong 250109, China*
[2] *School of Humanities, Arts and Education, Shandong Xiehe University, Jinan, Shandong 250109, China*
*Corresponding author. Email: huoxincun@163.com

ABSTRACT

Since the 18th century, *Tao Te Ching* has been translated into English over one hundred times. However, the research on these translations are far from enough in both quantity and quality, with few innovations in research method and perspective, which cannot match the significant status of the book all across the world. We try to construct a parallel corpus with the translation memory of Snowman and Live Docs of memoQ, which are regarded as popular computer-aided translation(CAT) software. In order to do that, we try choosing seven classical versions from the more than one hundred English versions based on the author, location and era. Some suggestions for possible future research are also given based on the functions of parallel corpora we have built.

Keywords: Tao Te Ching, Classical English Versions, Parallel Corpus

1. INTRODUCTION

As a Taoist classic in China, *Tao Te Ching* has been interpreted from the perspectives of philosophy, religion and human life. Across the world, it has been translated into a variety of languages, and the number of languages and versions of translation is second only to the Bible, known as the eastern bible. Given the strong ecclesiastic background of bible translation and the efforts of the large number of missionaries, *Tao Te Ching* could be the world's most translated book. According to incomplete statistics, as to English translations, there have been more than 100 versions in the past 100 years, among which the qualities vary. For the same sentence or even the same concept, different translators may adopt different methods and thus generate different translations based on the varied understanding. These varied translations not only confuse translation researchers and those who love and learn Chinese culture, and but also prevent the profound thoughts of Chinese culture from going out into the world as they really are. Therefore, it is of great significance to build a parallel corpus of *Tao Te Ching* by combining various English translations and the latest domestic research, which would be very helpful to the study of *Tao Te Ching* and Lao Zi, as well as the communication of Chinese culture.

Parallel corpus, as a statistical research method emerging in recent years, is one of the main tools for the quantitative study of literature, translation and other texts. With the help of corpus, we can explore the characteristics of roles, writing styles and even translation styles in our works. We can also build corpus of the same genre, such as news, laws, tourism, diplomacy, novels and poems. When we translate similar source text, we will have enough materials for reference. This is called fuzzy search in translation technology. In this paper, we will give a general introduction of *Tao Te Ching*'s English translation, overview the content, ideas and methods of corpus construction, and explore the roles that corpus can play in the study of the English translation of *Tao Te Ching*.

2. Summary of the English translation of *Tao Te Ching*

On the issue of the English translation of *Tao Te Ching*, some scholars have long discussed it at both macro and micro levels. Some will give a general investigation from the diachronic and synchronic point of view, while others focus on a term or a sentence to do a detailed study. Those who give a general view include the following scholars: Xin et al. [1] combed through the overall situation of the English translation of *Tao Te Ching* in the past hundred years, and proposed that its translation was closely related to the translation climax in the history of translation and the ideological trend of cultural exchanges between China and foreign countries. Zhang et al.[2] concluded that the existing translation of *Tao Te Ching*, like the translation of the Bible, should keep the Contention of a Hundred Schools of Thought by sorting out the number of English translations of *Tao Te Ching*. The studies of *Tao Te Ching* translation focusing on the details are as follows: Li [3] analyzed the core concepts and the metaphors in English expression; Yang [4] investigated the general situation of translation, understanding and acceptance; Bai [5] inquired into the meanings of "DE" at different levels. All these are important microcosmic studies. Based on previous studies, Wen et al. [6] proposed that the existing studies of *Tao Te Ching* mainly include the following aspects: the contents

unbalanced and concentrating on the first few chapters; focusing on the translated works by famous scholars, and insufficient comparative study and studies on the philosophical aspects of *Tao Te Ching*.

In view of the shortcomings of the comparative study in different versions, a similar phenomenon was found after carefully combing the articles related to *Tao Te Ching* published in the two top domestic translation journals Chinese Translators Journal and Shanghai Journal of Translators in recent years. When we use "*Tao Te Ching*" as the keyword to search articles in Chinese Translators Journal, there have been only six articles published in the last 30 years, and all in the last 10 years. Only one article is related to the construction of corpus, which was published by Zhao [7] in the 4th issue of 2015, namely, A Corpus-Based Study on the Translation Style of the Two English Versions of *Tao Te Ching* . This article has been downloaded 1396 times. However, there are only two articles about *Tao Te Ching* in Shanghai Journal of Translators in the past two decades, and one is The Study on English Translation of *Tao Te Ching* in China: Review and Prospect by Wen et al.[6] and the other is Analysis of Dao De Jing Translation by Wu Jingxiong from the Perspective of Metonymics by Zhao [8]. Then we used "Lao zi" as the keyword to search the same two journals, and found that there were only two articles in Chinese Translators Journal, which involved parallel corpus, including corpus-based translation teaching, corpus and lexicography. The Construction of Chinese-English Translation Parallel Corpus of Laozi published by Wang et al.[9] in the fourth issue of 2013 is the closest to the research object of this paper, which has been downloaded 681 times. He profoundly points out that there are 15-20 Chinese versions and more than 100 English versions in corpus construction. As a result, the upfront cost and time of corpus construction will be very demanding, which requires more people to cooperate and invest great efforts.

The above introduction basically reflects that while the national strategy of Chinese culture going out drives cultural studies, it also highlights that the quantity and quality of the studies on *Tao Te Ching* are obviously insufficient, and the research perspective lacks novelty. Moreover, it is rare to study *Tao Te Ching* by using parallel corpus. This is obviously not commensurate with the world-class status of *Tao Te Ching*. Therefore, the corpus-based study of different editions of translation really matters in understanding the true meaning of terms and accurately grasp philosophical ideas in *Tao Te Ching* in the style of translation, the translators' subjectivity and so on. We can also share lessons for similar classic research. Based on the above research, this paper aims to make up for the deficiency of the comparative study of *Tao Te Ching* through the construction of parallel corpus of multiple translation versions.

3. Construction of Parallel Corpus

Parallel corpus is one of the most important tools for quantitative research. Good tools are essential to do the job well and the construction of corpus is like the forging of tools, so its construction quality will determine directly the speed, quality and efficiency of subsequent research. It can also decide whether the corresponding research conclusions can be obtained in time. The success of building and using parallel corpus for relevant research depends on three factors: whether the content selection is reliable, whether the idea of building the database is correct, and whether the procedure of building the database is scientific.

3.1. Content Selection

Content selection, that is, the selection of translation plays a key role in the construction of parallel corpus. The typicality, the overall coverage of all kinds of translation, the quality of the translation and translators are all related to the successful construction of the parallel corpus, and determine whether it can effectively serve the subsequent research.

On the basis of the principle of inclusiveness, the author collected many English versions by authors from Britain, the United States, Japan, Hong Kong and mainland, and finally settled on seven versions over a period of more than 70 years: They were [10] James Legge (1891) and Arthur Waley (1934) in Britain, R.B. Blakney (1955) in the United States, D.T. Suzuki & Paul Carus(1880) in Japan, D. C. Lau in Hong Kong, and Lin Yutang, Xu Yuanchong and Gu Chengkun, all of whom famous scholars. They include famous sinologists from Europe and the United States, scholars from Japan who have a strong interest in Chinese culture, famous scholars with Chinese origin or professional translators from mainland.The seven translation covers a time span of nearly one hundred years, and their typicality and quality can basically reflect the general situations of *Tao Te Ching* translation, as well as the different understanding of the Chinese and foreign scholars. The content of *Tao Te Ching* laid a good foundation for parallel corpora construction and also provides a good material for subsequent research.

In the past three decades, other foreign scholars have been trying to translate *Tao Te Ching* into English, including [10] Stephen Mitchell (1988), Jane English & Gia-Fu (1989), Victor H. Mair (1990), S. Addiss & S. Lombardo (1993), J. H. McDonald (1996), and Ron Hogan (2004). However, on the whole, they are far less popular than the version the author chose and domestic translators paid little attention to it, so they are excluded from our list.

3.2. Construction Thinking

The corpus to be discussed in this paper belongs to parallel corpus, a one-to-many corpus, and one formed with the help of computer-aided translation(CAT) software, which is not substantially different from the traditional corpus. The only difference is that we don't achieve the alignment in the help of professional tools such as ParaConc. We just

make full use of the function of automatic segmentation in CAT software and then human intervention is imposed to achieve syntactic level alignment. This way, concordance function of CAT software can be used to search the related content with keywords, and get one and the same term translated with many more target expressions, thus achieving the purpose of using a variety of references. The corpus produced in this way is also known as raw corpus. After the sentence alignment is achieved, according to the research needs, the corpus can be further processed with professional morphological reduction and part-of-speech encoder tools such as AntConc and claws and further processed at grammatical and textual level. What I depend on is the function of corpus alignment in the CAT software memoQ and Snowman to preliminarily realize the alignment at the sentence level. Its original intention is to help readers at home and abroad to understand the classic book, and also facilitate the use of relevant researchers.

It is not my initiative to realize sentence level alignment, and segment and mark the corpus with professional corpus tools according to different research needs, which has been discussed by relevant scholars in China and across the world. As early as 2009, Wang et al.[11] proposed to build a bilingual parallel corpus by using translation memory in CAT software. The basic idea is to use translation memory to realize syntactic alignment first, and then use MS-Excel to mark the beginning and end of paragraph, the massive intelligent Chinese and English word segmentation system to achieve word segmentation, and then carry out the next corpus processing according to the research needs, such as error tagging and translation skills tagging. Finally, ParaConc, a bilingual parallel corpus retrieval tool, can be used to retrieve the marked corpus. This is an earlier proposal by domestic scholars to use CAT software as well as other professional tools to make corpus, which is of guiding significance to the construction of corpus.

3.3. Construction Process

Firstly, the author carries out preliminary processing to the text and removes the superfluous symbols and formats, so that the input text must be standardized and accurate. This process is called text cleaning and produces clean text. After that, the Chinese language was syncopated with the period as the basic unit based on the sentence meaning. The sentence segmentation is consistent with the corresponding English version. In the whole process of Chinese segmentation, it is necessary to ensure that the Chinese original text can consistently correspond to each sentence in the seven versions. Therefore, after the Chinese version is input into the corpus, the automatic segmentation of Chinese source text should be kept as much as possible. Based on this, the most inappropriate places are adjusted and marked so that subsequent English versions are consistent with this version. Secondly, we align the seven English versions with the same Chinese version. This lasted for a long time and realized the one-to-one correspondence between English and Chinese, that is, the seven English versions were aligned with each Chinese

sentence. This is the essence to build a one-to-many corpus. Through my own practice, such a bilingual aligned corpus can meet most of the needs of translation studies.

4. The Role of Parallel Corpus in Promoting the Relevant Studies of *Tao Te Ching*

Parallel corpus is based on the advantages of statistics, and has a very significant role in promoting traditional culture teaching and research and in researches based on a certain types of words (part of speech, appellation) or the overall translation style. Relevant scholars have made corresponding explorations in these aspects. By comparing the two versions of Wu Jingxiong(John C. H. W u, 1899-1986),a Chinese scholar and Arthur Waley (Arthur Waley) (1888-1966), a British Sinologist, Zhao [7] built a small corpus and got some interesting results by analyzing the two version with WordSmith 6.0 and self-edited software. The result in word is as follows:

Table 1 Statistical Results at Lexical Level

Item	Wu	Waley
Tokens	8,437	9,404
Types	1,817	1,774
Types/Tokens Ratio	21.54	19.03
Average word Length	4.26	4.27
Lexical density (%)	51.48	48.92
Common words(%)	70	62

She explained carefully the result from three perspectives of vocabulary, sentence and narrative perspective of the text, analyzes the translator's style, discusses the cognition, translation purpose and translation strategy behind it. This has important implications for our research, and further explains that the research method on the comparison of translator's style has made a break through in quantitative research. His research has achieved a combination of qualitative and quantitative research.

Based on the existing research, the author believes that parallel-corpus- based subsequent study can start from the following aspects:

(1) The linguistic features of the translation version must be further explored, which include words, phrases, sentences and texts, with many aspects such as parts of speech, sentence patterns and style included. For a classic such as *Tao Te Ching*, we can study the translation of a certain concept, that is, the study on terms, such as the comparative translation study of "the Tao" in seven versions. In addition, the advantages of quantitative analysis of corpus can be utilized to make a comparative analysis and research on the linguistic characteristics such as word frequency, word length, sentence length and

pattern and styles in different versions, so that we can find out the influence of the factors behind translation on the linguistic characteristics, and then infer the deep-rooted cultural motivation behind them.

(2)Based on the study of language features, we can search in depth the relationship between linguistic features and the native language of the translator and target readers, inquire into the culture's influence on the people in target language country in accepting an exotic culture, so as to explore and analyze the path and the laws of Chinese culture in spreading out to foreign countries and make corresponding contributions to the country's culture going out strategy.

5. Conclusion

There are few corpus-based studies on *Tao Te Ching*, but there are a lot of research findings on the other three of the Four Great Classical Novels, especially on *A Dream in Red Mansions*. Under the strong academic support of Redology, the corpus-based research on it can provide us with huge information and numerous novel perspectives. Therefore, we can get some inspirations from these perspectives and cutting points, and then apply them to the study of the *Tao Te Ching*, so that we can push such an amazing book that embodies the wisdom of the Chinese people and reflects the Chinese cultural tradition to its rightful position in the world.

ACKNOWLEDGMENT

REFERENCES

[1] H.J.XIN, S.B.GAO, Diachronic Description of *Tao Te Ching* in the English World, Journal of Nanjing Agricultural University (Social Sciences Edition), 2008(01),pp79-84

[2] X.G.Zhang, T.F. Bao, Classification and Review of the English Versions of Lao Zi' s *Tao Te Ching*, Journal of Jiangnan University(Humanities & Social Sciences) 2010,9(02), pp115-120.

[3] W.Z. Li, An Analysis of English Representation of the Key Concept and Its Metaphors in *Tao Te Ching*g, Journal of PLA University of Foreign Languages, 2015,38(05), pp108-116+150+160.

[4] H.L.Yang, A Case Study of "the Tao" in *Tao Te Ching*, Chinese Culture Research,2009,(03)pp192-196.

[5] J.R.Bai, C.Y.Yang, On the implications of the virtue in Daoism, Journal of Hebei Normal University(Philosophy and Social Sciences Edition),2015,38(06)pp121-125.

[6] J.Wen, Z.Luo, The Study on English Translation of *Tao Te Ching* in China: Review and Prospect, Shanghai Journal of Translators,2012,(01)pp19-23.

[7] Y.Zhao, A Corpus-based Study on the Translation Style of Two Versions of *Tao Te Ching*, Chinese Translators Journal,2015,36(04)pp110-113.

[8] Y.Zhao, Analysis of *Tao Te Ching* Translation by Wu Jingxiong from the Perspective of Metonymics, Shanghai Journal of Translators,2016,(03)pp56-61+94.

[9] D.M.Wang, Q.Y.Li, Construction of Parallel Corpus for Chinese-English Translation of Lao zi, Shanghai Journal of Translators,2013,(04)pp60-64.

[10]http://www.24en.com/subject/taoteching/translations.html

[11] Z.Wang, D.Y.Sun, Building Bilingual Parallel Corpora with Translation Memory Systems , Foreign Languages Research,2009,(05)pp80-85.