

Natural Language Processing-based Comparative Study of Translating Characteristics from the Version of *Tao Te Ching*

Hao Lin ^{1,a}, Zhonghua Han ², Li Yang ¹

¹ School of Foreign Languages, Shenyang Jianzhu University, Shenyang 110168, China;

² Faculty of Electrical and Control Engineering, Shenyang Jianzhu University, Shenyang 110168, China

^a 329547432@qq.com

Abstract. This paper established a small-scaled corpus based on modern corpus technology, applied corpus search software Antconc3.5.8 and SPSS 16 to analyze two translating versions of *Tao Te Ching* from John C.H.Wu and Arthur Wayley in terms of vocabulary, syntax and context. The paper further analyzes and illustrates reasons of these two translating versions from original text understanding, translating purpose understanding and translating strategy so as to provide references with translating style study of *Tao Te Ching*, translators' characteristics research and parallel corpus establishment.

Keywords: *Tao Te Ching*; translating characteristics; corpus research, comparative analysis.

1. Introduction

Tao Te Ching is one of the most important works in China and also an important source of Taoists' philosophical thoughts. It includes various subjects like philosophy, ethics, politics, military science, exerts profound influences on Chinese philosophy, science, politics, religion, etc, and reflects ancients' world outlook and outlook on life. With extremely high artistic value and artistic charm, *Tao Te Ching* is not just our national treasure in culture and art, but also the humans' spiritual wealth and cultural essence so many scholars at home and abroad are attracted to illustrate and translate. Along with many classical translation versions, *Tao Te Ching* is broadly and creatively developed in China even in the world so translating version research of *Tao Te Ching* can clarify sequence and direction in development of related theory for our national English translation of Chinese classics.

With deep development of English translation version of *Tao Te Ching*, a series of studies in translators' style, translating characteristics have been an important research topic to study English translation characteristics based on corpus. Many domestic scholars have broadly and comprehensively studied translating version characteristics of *Tao Te Ching*. For example, Xu Wentao comparatively analyzes self-established corpus of *Tao Te Ching* from three different English translation versions, discusses interactions between source language culture and target language culture in *Tao Te Ching* translation, and provide a totally new and multi-dimensional quantitative perspective for other classics translation. Wen Junchao discusses the importance of Chinese-English parallel corpus from theoretical perspective, analyzes status quo of domestic corpus establishment of *Tao Te Ching* corpus, and points out problems during corpus establishment. LV Wenpeng, LIU Hongru and YANG Xiangling study a self-built parallel corpus of *Tao Te Ching* and utilizes a combination of qualitative and quantitative methods. These two translation versions exhibit tendencies towards simplification and distinct translation characteristics but differ significantly in style. ZHANG Xuran, XING Yongle, ZHANG Pan and GE Lingling take translating version of *Tao Te Ching* as research objective, focuses on vocabulary and syntax to comparatively analyze translating language characteristics and explores the influence of translators' periods and identities on translating styles. Their studies discover that all translating versions have respective characteristics and most translators adopt direct translation method to keep Chinese culture element with simple words and refined sentences. In terms of translators' identities, western translators reflect longings for Asian culture but Chinese translators prefer to spread Chinese culture

and reflect translators' culture confidence. On the basis of above studies, translating version of *Tao Te Ching* has been studied by some scholars but it also needs comprehensive study from perspectives of vocabulary, syntax and semantics.

2. Corpus Selection and Parameters Setting

Among all English translating version of *Tao Te Ching*, one famous Chinese scholar John C. H. Wu and English Sinologist Arthur Waley all translate *Tao Te Ching*. Furthermore, their translating versions are usually considered as the most approximate to source text of *Tao Te Ching*. Thus, this paper will establish a small-scaled self-built corpus of these two English translation versions from these two translators. The reasons why this paper chooses these two translating versions are: First, since *Tao Te Ching* is one of important Chinese classics so its English translation version must be typical and representative; second, in terms of the published time, John C. H. Wu's translating version of *Tao Te Ching* was first published in 1939, which was in the journal of T'ien Hsia Monthly while Arthur Wayley's translating version was published in 1934 from Allen and Unwin company. On this basis, their translation versions were nearly published so there is comparison between these two translating versions. On the basis of above two important reasons, there is typical and representative for samples selection so it is suitable for comparative analysis.

3. Empirical Analysis of Corpus *Tao Te Ching* in Translating Characteristics

From vocabulary characteristics, syntax characteristics and discourse perspective, by means of corpus search software Antconc 3.5.8 and data analysis software SPSS 16, this paper studies translation characteristics of corpus based on above aspects for further data analysis. Data analysis includes types, tokens, type/token, average word length, vocabulary density, etc. Based on data analysis, translations from John C. H. Wu and Arthur Waley will be performed comprehensive corpus for further study.

3.1 Statistics and Analysis of Vocabulary

Corpus search software Antconc3.5.8 is applied to carry out data extraction from these two sub-corpus. On the basis of table 1, it is clear that Arthur Waley's translating version is 1000 more than that of John C. H. Wu in tokens but types of Arthur Waley's translating version are less than John C. H. Wu. Furthermore, John C. H. Wu's type/token is more than Arthur Wayley. This fully indicates that John C. H. Wu's translating version is less in repeating words but abundant in vocabulary but the density of John C. H. Wu is clearly lower than Arthur Wayley. All this proves that John C.H.Wu applies lots of functional vocabulary and his translating version is simple, clear, direct in expression so his translating version structure is more clear. Even though John C.H.Wu and Arthur Wayley are similar in average word length, difference in total number of 1-5 characters is obvious. This further proves that John C.H. Wu's translating version tends to use more common and strong function vocabulary than Arthur Wayley.

Table 1. Vocabulary Characteristics of Two Translating Versions

Version	tokens	types	Type/token	Total 1-5 characters	Average word length
John C.H. Wu	456	456	123	6330	4.25
Arthur Wayley	789	213	644	6164	4.28

In order to test whether there is difference in translating vocabulary between these two translating versions, data from software Antconc3.5.8 will be brought into statistical analysis software SPSS16. Since data are not in normal distribution, this test belongs to Chi-square test of non-parameter test.

In terms of Chi-Square test in SPSS, there valid value is 1.674, missing value is 0, total value is 1.674, Pearson Chi-Square value is 1.004, freedom value is 36, likelihood ratio is 3.368, linear

correlation value is 1.635, and quantity of effective data is 16.738. Bi-significant value is 0 which is smaller than key value of 0.05 so this proves that these two translating versions have significant difference.

3.2 Statistics and Analysis of Sentence Perspective

This paper will analyze translating sentences from sentence length and its total number since these two aspects reflect sentence complexity. By means of corpus search software Antconc 3.5.8 to extract data from two translating versions sentences, it is discovered that average sentence length of John C.H. Wu is 11.26 and total number of sentences is 750 while Arthur Wayley's average sentence length is 17.65 and total number of sentences is 529. On the basis of data, Arthur Wayley's sentence is longer and sentence structure is more complex. In order to test whether they have significant difference, data will be brought into SPSS16. Because data are not in normal distribution, SPSS 16 will be applied for Chi-square of non-parameter test.

Table 2. Sentence Characteristics of Two Translating Versions

Version	Average Sentence Length	Total Sentence Quantities
John C.H. Wu	11.26	750
Arthur Wayley	17.65	529

Based on above table 2, John C.H.Wu's mean is 3.81, standard deviation is 522.36, the minimal value is 11.26, the maximal value is 750 and Chi-square value is 2.798. Arthur Wayley's mean is 2.73, standard deviation is 361.58, the minimal value is 17.65, the maximal value is 529 and Chi-square value is 3.366. Freedom value of these two translating versions is 1 and significant value between these two translating versions is 0, which is smaller than 0.05. This indicates that their translations are in significant difference in terms of sentence. On the basis of data, Arthur Wayley applied lots of long sentences while John C.H. Wu applied relatively short sentences and sentence structure is more simple.

3.3 Statistics and Analysis of Context Perspective

Based on corpus search software Antconc 3.5.8, this paper will analyze and study context difference from narrative person, common conjunctions and prepositions. In terms of common conjunctions, John C.H.Wu's translating versions use 52050 common conjunctions while Arthur Wayley uses 42743 common conjunctions. In terms of common prepositions, John C.H.Wu uses 16540 common prepositions and Arthur Wayley uses 12600 common prepositions. It is clear that John C.H.Wu's common conjunctions frequency is obviously higher than Arthur Wayley but the frequency of common prepositions is less than the translating version of Arthur Wayley. There is no doubt that John C.H Wu's translating version is significantly different from Arthur Wayley.

Table 3. Common Conjunctions Distribution in Two Translating Versions of *Tao Te Ching*

Version	and	but	so	if	as	or	while	because
John C.H. Wu	28981	6080	3537	3884	377	2292	1499	798
Arthur Wayley	23366	4992	2897	3491	3451	1927	1137	669

Table 4. Common Prepositions Distribution in Two Translating Versions of *Tao Te Ching*

Version	apart from	due to	in spite of	because of	despite	rather than	thanks to	as a result of
John C.H. Wu	5981	3981	2547	2031	231	694	644	431
Arthur Wayley	4566	2897	2031	1547	197	587	531	244

4. Cause Analysis of Translating Characteristics Difference between Two Translating Versions

4.1 Understanding Difference of Original Text for Tao Te Ching

John C.H.Wu believes that *Tao Te Ching*, which represents Taoist culture, has common characteristics and homology with western culture, represented by christian so he focuses on similarity between these two translating versions during translating *Tao Te Ching*. Therefore, he is used to choosing simple, common, clear and functional vocabulary to efficiently transmit Chinese culture. However, Arthur Wayley believes that *Tao Te Ching* has Chinese culture particularity so he focuses on words accuracy and formal language usage. The reason is he makes efforts to more accurately understand Chinese culture rather than misunderstand Chinese culture for western readers.

4.2 Understanding Difference of Translating Purpose for Tao Te Ching

John C.H.Wu mainly transmits and interprets Chinese culture for western readers so he emphasizes acceptable degree of readers and focuses on relevance and interaction. Thus, he adopts simple and clear sentence pattern, transmits rich connotation of Chinese culture through simple words and bridges distance between readers and writers so as to reach western transmission of Chinese culture. However, Arthur Wayley's translation focuses on accuracy and makes efforts to clearly interpret each word meaning and present original text knowledge so he tends to use lots of clauses and complex sentence to explain specific vocabulary, which focuses on original meaning of writers.

4.3 Understanding Difference of Translating Strategy for Tao Te Ching

John C.H.Wu focuses on free translation and clearly express true meaning of original text so his translating version is simple as well as powerful for western readers to deeply and comprehensively understand *Tao Te Ching*. With relative free and flexible translation strategy, John C.H.Wu determines that he can use less tokens and and more types to transmit content of *Tao Te Ching* so that they can more easily understand its meaning. Arthur Wayley focuses on literal translation and original text expression so he tends to use long sentences and tokens in order to closely express initial structure, content and internal relationship of *Tao Te Ching* so his translation focuses on historical nature and ideological content.

5. Conclusion

Through empirical analysis of translating characteristics by corpus and study of translating style difference, there is significant difference on vocabulary and sentence pattern of translating versions from John C.H.Wu and Arthur Wayley. John C.H.Wu's translating version focuses on acceptable nature so he chooses simple, common and rich vocabulary, makes effort to create space and atmosphere of readers' communication so that readers will become easier to understand content of original text. However, Arthur Wayley focuses on original meaning of text and translating version abundance so he uses more accurate and loyal real words as well as complex sentence to present knowledge of original text as much as possible. Therefore, John C.H. Wu attaches importance to mutual similarity of translating version but Arthur Wayley focuses on historical function of translating version. This study provides reference for future corpus construction and corpus study of important vocabulary.

Acknowledgments

1.Applied Basic Research Program of Liaoning Provincial Science and Technology Department(2022JH2/101300253);2.Project of Liaoning Province Education Department of China

References

- [1] Lv Wenpeng, Liu Hongru, Yang Xiangling. A Corpus-Based Comparative Study of Translator's Style in Two English Versions of "Tao Te Ching". *Journal of Chengdu University of Technology(Social Sciences)*, 2023, 31(3): 91-99.
- [2] Lin Hao, Liu Fuhua, Wu Minghai. Translation Characteristics Study of Corpus-Based *Tao Te Ching*. *Journal of Shenyang Jianzhu University(Social Science)*, 2019,21(6):623-627.
- [3] Yao Zhenjun, Zheng Xuhong. An Application of a Self-Compiled Bilingual Domain Ontology into the Descriptive Criticism of English Translation of Tao Te Jing. *Journal of Dalian Nationalities University*, 2004, 16(2): 186-190.
- [4] Ma Jiaxin, Lv Changhong. A Study on "Xuan" of Tao Te Ching from the Perspective of Semantic Prosody: Analysis based on English Translation Corpus of Tao Te Ching and Corpus of Contemporary American English. *Journal of Southwest Jiaotong University*, 2019,20(1): 71-80.
- [5] Wang Xi. Gender Awareness in the Daodejing English translations from the Perspective of Corpus-based Critical Translation Studies: A Case Study of the Explication of Personal Pronoun. *Journal of Chongqing University of Technology(Social Science)*, 2018, 32(4): 142-152.