

A Corpus-based Analysis on the Ancient Chinese Costume Terms in English

Zhou Dan^{1,a}, Chen Yonglin^{2,b,*}

¹School of Foreign Languages, Wuhan Textile University, Wuhan, 430200, China

²Hubei Preschool Teachers College, Wuhan, 430223, China

^aWindy_zhou@163.com, ^bchenyonglin693@163.com

*Corresponding author

Abstract: This study embarks on constructing a Chinese-English bilingual corpus dedicated to the systematic collection, organization, and analysis of costume terms found in classical Chinese literature, including "The Book of Songs," "Chu Ci," and "A New Account of Tales of the World." Aiming to bridge cultural and contextual gaps between Chinese and English, the research delves into effective extraction methods for costume terms, addressing linguistic accuracy and cultural adaptability. Through meticulous analysis of ancient costume terms, the study aspires to enhance the global dissemination of Chinese costume culture and improve the quality of cross-cultural communication. Employing Natural Language Processing (NLP) tools, the research elucidates the process of data collection, terms extraction, and the subsequent construction of a parallel corpus for comparative analysis. This endeavor not only illuminates the historical and cultural essences embedded in Chinese costume culture but also offers fresh insights into corpus-based linguistic theory and practice.

Keywords: Costume Terms; Chinese-English Bilingual Corpus; NLP

1. Introduction

Ancient Chinese costume, as an important part of Chinese culture, not only carries rich historical and cultural values, but also attracts wide attention internationally. This study aims to construct a Chinese-English bilingual corpus in order to systematically collect, organise and translate the Costume terms in classical Chinese literature such as *The Book of Songs*, *Chu Ci*, and *A New Account of Tales of the World*, in order to explore the historical and cultural values of Chinese costume culture. The study will focus on how to effectively extract costume terms from classical literature, deal with cultural and contextual differences between Chinese and English, and how to assess the accuracy and cultural adaptability of the rendition.^[1-3] Through in-depth analysis and accurate rendition of ancient costume terms, this study is expected to promote the international dissemination of Chinese culture and improve the quality of cross-cultural communication, as well as provide new perspectives and references for corpus-based linguistic theory and practice.

2. Corpus Construction

This paper will discuss in detail the process of data collection and processing, which is divided into three subsections covering data source selection and collection, terms extraction and organisation, and the application of NLP tools.

2.1 Data Source Selection and Acquisition

Determining the data source is the first task in building a corpus. In this paper, literary works representing the peak of ancient Chinese literature and different historical periods were chosen as data sources. The following is a detailed description of the data sources:

2.1.1 Selection of Literary Works

Our research data sources include the best of ancient Chinese literature, covering literary works from different historical periods and cultural backgrounds. These literary works include *The Book of Songs*, (reflecting early Chinese culture and costume), *Chu Ci* (a representative of Chu literature, rich in cultural

elements and depictions of costume), and *A New Account of Tales of the World* (reflecting the culture of the Wei and Jin periods and rich in costume terms).^[4] The selection of these works aims to delve into ancient costume terms while reflecting Costume culture in different historical and cultural contexts.

2.1.2 Data Collection Methods

Data collection is a crucial step in the study, and the following key steps were carefully planned to ensure the accuracy and completeness of acquiring primary documents. Firstly, the original texts of the aforementioned literature were acquired through a variety of means, including visits to libraries, digital libraries, and online resources.^[5] Text screening was carried out to ensure that the selected texts were highly relevant and representative of the research objectives.^[6] During the screening process, non-essential sections of text were excluded to minimize potential noise and improve data quality to ensure that the final data source met the requirements of the study.

2.2 Terms extraction and organisation

In the key stage of data collection and processing, terms extraction and collation is an important task in the study, aiming at accurately extracting and collating ancient Costume terms from the primary literature. The detailed steps of terms extraction are described below:

2.2.1 Terms extraction

Analysis of the primary literature: This paper begins with a detailed analysis of each literary work, carefully reading the text to identify passages and sentences that may contain costume terms. This step required an in-depth understanding of the content and context of the literary work in order to ensure accurate extraction of costume-related terms.

Identification of terms: the process of identifying potential terms involved analyzing and tagging the text with the help of natural language processing tools and terms dictionaries. These tools help us to automatically identify potential terms and provide the basis for subsequent validation and collation.

Validation of terms: After automated extraction, it performed a term validation process to ensure that the extracted terms are apparel-relevant and accurate. This step requires human intervention to exclude possible misidentified or irrelevant terms to ensure the quality and accuracy of the final term data.

Through the above steps, this paper will accurately extract terms related to ancient costumes from the original documents, which provides a reliable database for the subsequent corpus construction and comparative analysis of renditions. These terms will be collated into an exhaustive glossary of ancient Costume terms to support the objectives of the study.

2.2.2 Glossary of Ancient Costume Terms

The creation of a glossary of ancient costume terms is a crucial step in ensuring the neatness and usability of the research data, and is intended to provide a comprehensive understanding of ancient costume terms and a basis for subsequent comparative rendition analyses. This glossary will include the following key elements:

Name of the term: The glossary will record the name of each ancient costume term to ensure clear definition and identification of the term.

Definition of the term: Each term will be accompanied by its detailed definition to provide an accurate understanding of the term. This will help researchers and translators clarify the meaning and usage of the term.

Sources of terms: This paper will document the source of each term, i.e., the specific literary work from which the term was taken. This helps in tracking the literary context of the term and provides the research with citations and references to the literary work.

Examples of terms in context in literary works: In order to better understand the context and usage of terms, this paper will provide examples of each term in context in literary works. These examples will highlight the application of the terms in different contexts and will help translators and researchers to better understand the polysemy of terms and differences in usage.

This Glossary of Ancient Costume Terms will serve as a central resource for research, providing essential reference and support for rendition work. By exhaustively documenting and defining each term, we will ensure that the data for the study is neat and easy to use, as well as providing a powerful tool for

comparative rendition analyses to ensure accurate renditions and cultural adaptation of costume terms.

2.3 Application of NLP Tools

NLP tools play a key role in data processing, including text pre-processing and automatic term extraction. The application aspects of NLP tools are discussed below:

2.3.1 Text Preprocessing

The application of NLP tools in this study plays a key role, especially in the pre-processing stage of text data. Text pre-processing is a key step in ensuring the quality and consistency of textual data, with the goal of transforming raw documents into clean, consistent and structured data that can be subsequently processed. The following is a detailed description of the application of NLP tools for text preprocessing:

Noise removal: In text preprocessing, the first task is to remove noise from the text, which includes punctuation, special characters, numbers, and other non-critical information. The purpose of this step is to clean up the text and make it easier to analyse and process. NLP tools can be used to automatically identify and remove this noise, improving the quality and consistency of the data.

Text standardization: Text standardization is the conversion of text into a consistent format to ensure consistency in subsequent processing. This includes standardizing case in text, dealing with acronyms and synonyms, and removing formatting differences in text. NLP tools can be used to automate these standardization tasks to ensure text consistency, reducing confusion and errors when analyzing data.

Lexical annotation: Lexical annotation is the task of lexically classifying words in text. In this study, lexical annotation is very important for identifying and extracting terms. The NLP tool automatically assigns lexical labels to each word in the text, such as nouns, verbs, adjectives, and so on. This helps to identify key terms in the text and provides important information in the subsequent automatic term extraction process.

Overall, the application of NLP tools in text pre-processing helps to optimize the raw documents and make them suitable for subsequent data analysis and term extraction. These tools improved the quality, consistency and processing efficiency of the data, providing a reliable data base for the study and helping to achieve the research objectives of this study.

2.3.2 Automatic extraction of terms

Among the key stages of data collection and processing, automatic extraction of terms is a crucial step that utilizes text mining techniques and NLP tools with the aim of reducing manual workload and improving the efficiency and accuracy of data processing. The following is a detailed description of automated term extraction:

Term Extraction: This step uses NLP techniques to automatically extract possible terms from the text. The NLP model is able to identify specific terms in the text, such as ancient Costume terms, by analyzing the words, phrases and contextual information in the text. The extracted terms may include names of ancient costumes, specific styles, fabrics, costume parts, etc. This automatic extraction process is based on the understanding of language models and statistical analyses, and can quickly capture a large number of terms.

Term Filtering: The list of automatically extracted terms may include some irrelevant words or misidentified terms. Therefore, after the terms are automatically extracted, they need to be filtered and validated. This process involves filtering out terms that are relevant to ancient Costume while excluding those that are not relevant to the study. Criteria for filtering can include contextual information about the term, definition of the term, or keywords to ensure that the extracted terms are consistent with the research objectives.

The automatic extraction and filtering of terms enables researchers to quickly access a large number of ancient Costume terms and reduces the workload of manual collation and filtering. These automatically extracted terms will become important resources for subsequent corpus construction and rendition comparison analysis, providing a solid foundation for the research. Meanwhile, the application of NLP tools plays a key role in improving the efficiency and accuracy of data processing.

3. Construction of the Parallel Corpus

3.1 Corpus structure and management

To construct a well-organized and easily manageable corpus, this paper adopts a series of steps to organize and archive ancient Chinese literary works and their English translations. First, three selected ancient literary works are processed, with each work having a dedicated folder to store the original text and two different English translations. These translations represent different translation strategies and cultural contexts. Each folder contains the original Chinese text, English Translation 1, and English Translation 2. This method enables researchers to systematically archive and organize these literary works, making the comparison and analysis between the original text and the translations more convenient and clear.

In the process of constructing the parallel corpus, the original text and the two English translations were aligned on a sentence-by-sentence or paragraph-by-paragraph basis. This alignment ensures the comparability of the texts, laying a foundation for subsequent analysis of translation strategies and terminology studies. This parallel alignment method allows researchers to directly compare how the same paragraph or sentence is presented in different translations, offering deeper insights into the strategies adopted by translators, especially regarding the translation of cultural terms.

At the same time, each sentence or paragraph in the corpus has been carefully annotated to distinguish the source of the different works and translations. This data annotation method allows researchers to track the context and origin of each term, helping them understand the usage of terms in different contexts. Additionally, to ensure traceability during the data processing, this paper utilizes a version control system to document different versions of the corpus. This system allows researchers to revert to earlier states of the corpus at any time, enabling comparisons of the corpus at different stages. This system provides a safeguard for updating and adjusting the corpus, making data management and analysis more organized and efficient.

3.2 Comparative analysis of literary works and their two different English renditions

Before making a comparison of the translations, it is necessary to select representative sentences from the original text, especially those containing cultural terms. These terms not only literally relate to ancient Chinese dress, but are also deeply rooted in the social class, ritual system and aesthetic standards of the time. For example, the dress terms in the *The Book of Songs* may not only refer to one type of costume, but they may also symbolise multiple meanings such as social status, gender roles, religious beliefs and so on. Therefore, when translators deal with such terms, they need to consider their literal meanings as well as convey the cultural connotations behind them. By choosing these culturally connotative sentences as the object of analysis, we focus on the different translation strategies of translators when facing culturally loaded terms.

By analysing the ways in which different translators deal with the ancient dress terms in the translation process, we are not only able to gain an in-depth understanding of the translators' translation strategies, but also able to explore the complexity and diversity of translation as a tool for cross-cultural communication. Different translation methods not only reflect the translators' personal translation style and cultural understanding, but also reveal the inevitable cultural variation and adaptation phenomenon in the translation process. Ultimately, such comparative analyses help us to better understand the ways in which cultural terms are transformed in different languages and cultures, and provide theoretical support and practical reference for future translation practices.

3.3 Extraction of Sentence Pairs

Extracting sentence pairs containing specific ancient costume terms is a crucial step in this study, as these pairs will be used to analyze the usage and translation strategies of such terms in different renditions. The extraction method begins with selecting sentence pairs through text analysis and keyword matching. For instance, if the focus is on the term "dragon robe," sentences containing this specific term will be identified in both the original text and its two translations. This systematic approach ensures that the selected sentences are relevant and directly related to the costume terms being studied, allowing researchers to conduct focused comparisons between the different translations.

Once the sentence pairs are extracted, they will be tagged to indicate their source and the specific costume term they contain. This tagging process not only provides a clear reference for the context and

origin of the terms but also facilitates the organization of data for further analysis. By examining the differences in the translation of these specific ancient costume terms, the study can explore the nuances of rendition strategies in greater detail. This method allows for a more comprehensive understanding of how cultural terms are handled across different translations, offering valuable data for the classification and categorization of translation approaches.

4. Classification and categorisation of translated terms

Classification and categorisation of costume terms is a crucial part of cross-cultural translation, especially challenging when it comes to ancient costume culture. First of all, the categorisation of terms should be based on the type of dress, the historical period of its use and its cultural characteristics. In the process of translation, translators need to fully consider the cultural connotations behind the terms, especially when the terms are not limited to literal meanings, but also carry rich historical, social and religious information. For example, dress terms from different historical periods may symbolise the social class, ritual system and other specific cultural backgrounds at that time, so the translator needs to understand the multiple meanings of the terms in the original text through the context, so as to ensure that the translated text can be effectively conveyed in different cultures. The purpose of categorisation is to help the translator identify the context and usage of the terms and express them reasonably in relation to the culture of the target language, so that the translated text can accurately convey the original meaning and fit in with the cultural background of the readers at the same time.

In the process of categorisation, translators often need to further subdivide the translations of these dress terms to ensure that each term can be translated in a way that both maintains cultural fidelity and ensures the readability of the translation. When dealing with these terms, translators may choose a variety of approaches such as transliteration, direct translation, or paraphrase, depending on the cultural specificity of the terms and the receptivity of the target readers. Phonetic translation can help to preserve the form and cultural characteristics of the original text, while paraphrasing focuses more on conveying the cultural meaning of the terms to the target language readers. The process of categorisation not only helps to systematically analyse the translation strategies of different translators, but also reveals the cultural differences and linguistic adaptations in terms translation. This classification and categorisation method not only helps researchers to understand cultural transmission in translation, but also provides an important basis for further analysis of different translators' tendencies in dealing with culturally loaded terms.

5. Conclusion

This paper thoroughly analyzed the rendition strategies of costume-related terms in the ancient literary works *The Book of Songs*, *Chu Ci*, and *A New Account of Tales of the World*, and examined the impact of these strategies on cultural adaptation.^[7] The analysis revealed that translators employ various strategies, including naturalization and alienation, when translating costume terms, and that these strategies significantly influence the balance between comprehensibility and cultural fidelity in the translated texts. The naturalization strategy tends to make the text more accessible and understandable to English readers, while the alienation strategy focuses on preserving the unique cultural features of the original text, maintaining its authenticity.

The study also highlighted the importance of these strategies in addressing the challenges of cross-cultural rendition. Naturalization helps bridge cultural gaps by adapting foreign concepts into more familiar ones for the target audience, while alienation retains the original cultural nuances, offering readers a more immersive experience of the source culture. This balancing act between making a text relatable and preserving its original cultural essence is a core issue in literary translation.

Future research directions proposed in the paper include further exploration of the relationship between the choice of rendition strategies and the cultural background of target readers, as well as how to better strike a balance between the acceptability and originality of translated texts. Additionally, investigating the translation of costume-related terms across different literary works could provide deeper insights into cross-cultural rendition practices. These studies are expected to contribute valuable knowledge to the field of intercultural communication, fostering greater understanding and interaction between different cultures, and offering practical guidance for the global dissemination of literary works.

Acknowledgements

*This paper was sponsored by Ministry of Education Humanities and Social Sciences Fund Youth Fund Project "Research on the English Translation of Costume Terms in Chinese Classical Literary Works". (Project No. 21YJC740083) and the research project of Philosophy and Social Science of Hubei Provincial Department of Education "Research on the Construction of a Standardized Chinese-English Corpus for Beijing Opera Costumes" (Project No. 20Y073).

References

- [1] Zhou Dan, Yu Xuan, and Tan Yanbao. 2021. *Translation of Chinese Costume culture*. Wuhan: Wuhan University Press.
- [2] Schleiermacher A. 1992. *On the different methods of translating*. In R. Schulte & J. Biguenet (eds.). *Theories of Translation: An Anthology of Essays from Dryden to Derrida*. Chicago & London: The University of Chicago Press.42.
- [3] Nida E. A. 1982. *Translating Meaning*. San Dimas: English Language Institute.
- [4] Shang X. 2014. *English-Chinese Tense Style Types and Translation Strategies*. Shanghai: Shanghai People's Publishing House.
- [5] Wang Xiuwen. 2020. *A corpus-based study on the translation of ancient Chinese philosophical terms*. *Journal of Foreign Languages* (1):109-114.
- [6] Meng Lan. 2021. *A study on the foreign dissemination of traditional Chinese ideological and cultural terms in English translation*. *Shanghai Translation* (3):56-60.
- [7] Xu Mingwu, Zhao Chunlong. 2018. *The name and reality of domestic corpus translation studies*. *Shanghai Translation*, (4): 3-9.