Academic Education
Publishing House
-AEPH-

A Quantitative Study of the Chinese Original and English Translation of Alai's the Song of King Gesar

Yang Du, Xiaojin Zhang*

School of Foreign Languages, North Minzu University, Yinchuan, Ningxia, China
*Corresponding Author

Abstract: This study takes the Chinese original of "the Song of King Gesar" written by Alai and its English translation by Howard Goldblatt and Lin Lijun as the research objects. Using the research methods of quantitative linguistics, it conducts a comparative analysis of the Chinese original and the English translation from three dimensions: lexical richness, word frequency, and lexical complexity. The study aims to reveal the similarities and differences between the translated version and the original in terms of vocabulary, explore the adjustments made by translators in maintaining the original style and adapting to target language readers during the translation process, and provide references for the translation and dissemination of national epics. The study finds that the English translation is slightly lower in lexical richness than the Chinese original, shows certain differences in word frequency distribution, and makes appropriate adjustments in lexical complexity according to the characteristics of the target language. These results have important implications for understanding the translation rules of national epics and improving translation quality.

Keywords: Alai's the Song of King Gesar; Chinese-English Comparative Study; Quantitative Research; Lexical Richness; Word Frequency; Lexical Complexity

1. Introduction

1.1 Research Background

Ethnic epics are the cultural treasures of a nation, carrying its historical memories, values, and aesthetic tastes. The Epic of King Gesar, created by the Xizang people of China, is one of the longest epics in the world, possessing immense literary, historical, and cultural value. Centered on the heroic deeds of King Gesar, it depicts various aspects of Xizang society and serves as crucial material for studying Xizang culture.

As a renowned contemporary writer, Alai has reinterpreted the epic "the Song of King Gesar" through his unique perspective and narrative style. While preserving the core spirit and main plot of the original work, his revised version incorporates modern storytelling techniques and contemporary perspectives, making it more accessible to modern readers. This adaptation has expanded the audience for "the Song of King Gesar" and enhanced its influence in the contemporary literary world. Alai's reinterpretation not only preserves and innovates ethnic cultural heritage, but also pioneers new pathways for modern dissemination of national epics.

With the deepening development of globalization, cultural exchanges have become increasingly frequent, and the translation and dissemination of national epics have become an important bridge for promoting mutual understanding among different cultures. The English translation of "the Song of King Gesar" by Lin Lijun and Howard Goldblatt introduced this great national epic to readers in the English-speaking world, making significant contributions to the international dissemination of "the Song of King Gesar". As a renowned sinologist and translator, Howard Goldblatt has rich experience and profound expertise in translating Chinese literary works into English. The translation he collaborated with Lin Lijun has had a wide influence in the English-speaking world and is of great significance in promoting the Chinese national culture to the world.

1.2 Research Objectives and Significance

This study employs quantitative linguistic methods to conduct a comparative analysis between the Chinese original version of "the Song of King Gesar" as written by Alai and its English translations by Howard Goldblatt and

Lin Lijun. The research quantitatively examines lexical characteristics at the vocabulary level, revealing similarities and differences in translation practices. Specifically, by analyzing indicators such as lexical richness, word frequency, and lexical complexity, the study investigates the extent to which translators reproduce the original's lexical features during translation and the adjustments made to accommodate target language readers.

This study employs the method of quantitative linguistics to investigate the translation of ethnic epics, enriching the methodologies and perspectives of translation research. Through quantitative analysis, it can more objectively and accurately reveal the patterns and characteristics in the translation process, providing empirical support for the development of translation theory. The research findings can serve as a reference for the translation practice of ethnic epics, helping translators better grasp the lexical features of the original work, make more reasonable decisions during the translation process, and improve translation quality. At the same time, it also helps readers better understand the relationship between the translated text and the original work, enhancing their appreciation of ethnic epics. This study contributes to the international dissemination of "the Song of King Gesar" and promotes cultural exchange and integration. By studying the translated version, it can identify problems and shortcomings in the translation process, providing a basis for further improving translation work and enhancing the international influence of China's ethnic culture.

1.3 Structure of the Thesis

This thesis is divided into five sections. The first section serves as an introduction, outlining the research background, objectives, significance, and overall structure of the study. The second section presents a literature review, synthesizing domestic and international research on "the Song of King Gesar" including studies of the original text, Alai's retelling version, and its translations. It also examines the application of quantitative linguistics in translation style analysis to provide theoretical foundations. The third section details the research methodology, specifying corpus selection criteria and processing techniques for Chinese original texts and English translations—including data cleaning, noise reduction, and lexical annotation.

The fourth section presents results and discussions, analyzing the Chinese original and English translation versions through three dimensions: lexical richness, word frequency, and lexical complexity. The fifth section concludes with findings, limitations, future research directions, and implications for the translation and dissemination of ethnic epics.

2. Literature Review

As a core carrier of Xizang culture, King Gesar has been studied in multiple fields such as history, literature, and anthropology. The following sorts out the research achievements from three dimensions: studies on the original epic, studies on Alai's retold version, and translation studies.

Li (2011) observed that three distinct styles—Yang Xianyi's free-verse translations, Xu Yuanchong's rhymed translations, and Weng Xianliang's prose translations—have all been adopted in the English renditions of Li Qingzhao's ci-poems[1]. Qin and Wang compared the high-frequency word distribution between original Chinese texts and their English translations in their work A Corpus-Based Study on the Lexical Features of Translated Texts. They found that translators tend to use more stable core vocabulary to reduce the cognitive load of target language readers[2]. Zhang pointed out in A Study of Howard Goldblatt's Translation Style — Taking the English Translation of King Gesar as an Example that the translation by Howard Goldblatt and Sylvia Lin adopted a "dynamic equivalence" strategy[3]. Yang and Yang defined the highest level of respect for the original work as fidelity to its literary style[4]. Liu and Fang (2017) conducted a comparative study on the variations in translator style among the Chinese versions of Stray Birds through qualitative and quantitative methods, and the research indicated that Feng Tang's translation exhibits significant stylistic deviation[5]. Zhang pointed out in her article Is Chinese "Hua'er" Still "Hua'er" in the English-Speaking World? that the stylistic features of "Hua'er" are closest to those in the self-built Chinese folk song corpus, followed by modern Chinese poetry[6]. Yang studies Epic of Gesar of Ling's verse via English ballads, noting their shared traits, analyzing the epic's verse features, and stating the ballad-based approach aids minority epic translation[7]. Li and Zhang (2024) introduces quantitative linguistic methods

into the research on the Yi creation epic Hnewo Teyy. Through word frequency analysis, Zipf's Law verification, and clustering experiments, it reveals the epic's narrative style characteristics, fills the gap in relevant research methods, and contributes to the systematization of ethnic epic research[8]. Xiang introduces quantitative methods to explore the textual characteristics of the Tujia epic Sheba Songs. Through word frequency verification, part-of-speech distribution analysis, and clustering experiments, it reveals the epic's proximity to modern poetry and tendency toward folk balladization, addressing the limitations of previous qualitative studies[9]. Danzhu Angben emphasizes protecting its spiritual value, culture, artists and disciplinary system, highlighting its significance for Chinese and human civilization[10].

In conclusion, existing studies have provided rich perspectives on the textual characteristics and translation strategies of King Gesar, yet a systematic quantitative analysis of Alai's version and its English translation remains a gap. This study will integrate the aforementioned theories and methods, and attempt to address this aspect through quantitative comparison at the lexical level.

3. Research Design

3.1 Corpus Selection
The Chinese original corpus for this study is "The Epic of King Gesar" by Alai, a reinterpreted version based on the traditional text. This work, containing approximately 216,441 characters, holds significant representativeness and influence. The English translation adopted is the version translated by Howard Goldblatt and Lin Lijun, comprising about 113,784 characters, which has gained widespread recognition and dissemination in the English-speaking world.

The selected Chinese original works and English translations were cleaned to remove irrelevant information such as page numbers, comments, and copyright information to ensure the purity of the corpus.

Clean the text by removing noise data such as typos and punctuation errors to enhance corpus quality. For Chinese texts, perform word segmentation to split continuous Chinese character sequences into meaningful words. For English texts, apply word form normalization to standardize verb tenses and noun singular/plural

forms into their original forms.

The processed corpus undergoes lexical annotation, including part-of-speech tagging. Chinese lexical annotation follows the Chinese part-of-speech tagging system, while English lexical annotation adopts the English part-of-speech tagging system, providing a foundation for subsequent lexical analysis.

3.2 Research Methods
3.2.1 Lexical richness
Lexical richness is a key metric for evaluating textual diversity. This study employs the Type-Token Ratio (TTR) to measure lexical richness, where Type denotes the frequency of different lexical items and Token represents the total lexical count. A higher TTR value indicates greater lexical richness. The calculation formula is: TTR= Type Count / Token Count. To ensure a more comprehensive assessment, the standardized Type-Token Ratio (STTR) is calculated to eliminate the influence of text length on TTR.

3.2.2 Word frequency
Word frequency refers to the number of times of a specific word within a text. This study will conduct frequency analysis on vocabulary in both the original Chinese text and its English translation, examining the distribution patterns of high-frequency words. By comparing these high-frequency words, we will explore similarities and differences in vocabulary selection between the translation and the original work. Additionally, we will analyze the frequency distribution of specific word types (such as nouns, verbs, adjectives, etc.) to reveal distinctive characteristics in their lexical usage.

3.2.3 Lexical complexity
Primarily encompasses lexical length and lexical difficulty. This study measures lexical complexity through two dimensions: lexical length and lexical difficulty. Lexical length is reflected by calculating the average word length, defined as the ratio of total characters to total words. Lexical difficulty is assessed by referencing a lexical difficulty scale, which quantifies the proportion of words at different difficulty levels within the text. This methodology enables analysis of lexical complexity differences between Chinese original works and their English translations.

4. Results and Discussion

4.1 Lexical Richness

4.1.1 Chinese original work

Analysis of the Chinese original work of Alai's "The Song of King Gesar" reveals a lexical richness ratio (TTR) of 0.11 and a standardized lexical richness ratio (STTR) of 0.092. This indicates that the original Chinese text demonstrates notable lexical diversity, with the author employing a rich vocabulary to vividly depict narrative plots and character portrayals, thereby enriching the textual content. From a grammatical perspective, different parts of speech exhibit varying numbers of lexical categories and forms, with nouns and verbs showing relatively higher category counts, suggesting the original work emphasizes descriptive narration of objects and actions.

4.1.2 English translation

The English translation demonstrates a TTR of 0.07 and STTR of 0.05, slightly inferior to the Chinese original. This suggests the English version exhibits lower lexical richness. The disparity may stem from multiple factors: Firstly, as distinct language systems, Chinese and English differ in lexical expressions. To align with English conventions, translators might prioritize common vocabulary, potentially reducing lexical richness. Secondly, translators must consider target readers' comprehension levels, often simplifying complex or obscure terms to enhance readability. These considerations collectively contribute to the observed differences in lexical richness between the Chinese original and its English translation. This difference reflects the characteristics of the two languages and the considerations of the translator in the translation process. The original Chinese work shows its literary charm through rich vocabulary, while the English translation pays more attention to the readability and understandability of the translation while maintaining the meaning of the original work.

4.2 Word Frequency

4.2.1 Original Chinese text

Following the Table 1, "Gesar"ranks first with a high frequency of 1,063 times, far exceeding other words, directly highlighting the epic's narrative logic centered on the hero. As the core figure in Xizang epics who subdues demons and benefits the people, the frequent appearance of his name reinforces the protagonist's aura and aligns with the traditional epic feature of connecting events through characters.

Table 1. High-Frequency Words in the Original Chinese Version

Rank	Type	Frequency
1	Gesar	1063
2	Ling Kindom	425
3	Guo'wang	376
4	minister	176
5	steward	172
6	hero	156
7	Da'wang	154
8	horse racing	98
9	defeat	12
10	Lama	128

The repeated use of "Guo'wang" (376 times) and "Da'wang" (154 times) not only clarifies Gesar's identity but also implies the epic's themes of royal power and responsibility — Gesar's heroic deeds reflect not only personal bravery but also a close bond with the rise and fall of "Ling Kindom" (425 times) and the well-being of its people. The high frequency of "hero" (156 times) directly defines the group trait of Gesar and his followers; through lexical repetition, it strengthens the epic's core theme of celebrating heroism.

"Ling Kindom" (425 times), as the core region of Gesar's activities, its frequent appearance constructs a specific narrative space for the epic. It allows readers to clearly perceive the story's setting and implicitly carries the symbolic meaning of "Ling Kindom"(425 times) as the spiritual homeland of the Xizang ancestors.

Words such as "minister" (176 times) and "steward" (172 times) outline the political structure of "Ling Kindom"(425 times), reflecting the power distribution model of ancient Xizang society and adding historical authenticity and social depth to the epic. The repeated appearance of these words not only lays the groundwork for plot scenes like consultations and decision-making but also implicitly conveys the organizational culture of Xizang society.

The high frequency of "Lama" (128 times) is a direct projection of Xizang religious culture in the text. As an important role in Xizang culture, "Lama"often performs the functions of guidance and blessing in the epic. Their frequent appearance not only conforms to the historical background of the integration of religion and politics in Xizang society but also endows the epic with sacred and spiritual connotations. Although "horse racing" (98 times) has a lower frequency than core character-related words, it

holds iconic significance — in the traditional narrative of King Gesar, winning the horse race is a key event for Gesar to gain royal power. The repeated use of this word emphasizes the turning point of fate in the epic, enhancing the dramatic tension of the plot. "defeat" (12 times) has a relatively low frequency, but combined with the epic's main line of subduing demons and resisting foreign invaders, it is associated with the narrative of numerous war scenes. The restrained use of this word, instead, highlights the preciousness of victory and Gesar's brave qualities through a concise yet refined approach.

4.2.2 English translation

There is a certain corresponding relationship between the high-frequency words in the English translation and the original Chinese text. However, there are some differences in word frequency distribution. For instance, some culture-loaded words that appear frequently in the original Chinese text have relatively low frequencies in the English translation. This may be because the translator adopted free translation or explanation for these words during the translation process to facilitate the understanding of target language readers. Meanwhile, some conjunctions and prepositions in the English translation have relatively high frequencies, which is determined by the grammatical characteristics of English — connections between English sentences rely more on conjunctions and prepositions.

Following the Table 2, "King" tops the list with 740 times, surpassing "Gesar" (552 times). This change implies the narrative habits of the English-speaking world — in the Western epic tradition, "king" is an identity symbol more easily understood by readers. By increasing its frequency, the translator lowers the cognitive threshold of the target readers for "Gesar", an exotic hero. The retention of "Gesar" ensures the recognizability of the core character's name, achieving a balance between "exoticism" and "comprehensibility". "Tribe" (70 times) corresponds to it in Chinese but has a lower frequency than the relevant words in the original text, which is related to the weakened concept of "tribe" in English society. In modern English context, "tribe" often carries implicit connotations of "primitiveness and marginality". By reducing its frequency, the translator avoids misinterpretation of Lingguo's society and better aligns with English readers' familiarity with the "kingdom" system.

Table 2. High-Frequency Words in the English Translation

Rank	Type	Frequency
1	king	740
2	gesar	552
3	horse	207
4	heaven	187
5	tribe	70
6	fight	36
7	Demon	139
8	hero	37
9	war	38
10	princess	38

The high frequency of "heaven" (187 times) is a translation of the "divine" elements in the original text. Expressions related to "heaven and gods" in the original Chinese text are relatively scattered. The English translation integrates these scattered sacred images into a unified cultural symbol through the concentrated use of "heaven", which not only simplifies the expression but also conforms to the linguistic habit of "heaven" as a core sacred concept in English. "Demon" (139 times) corresponds to the "demon" related words in the original text, and its high frequency strengthens the narrative conflict of "opposition between good and evil". In the Western literary tradition, "demon" is a classic villain symbol. By increasing its frequency, the translator amplifies the dramatic tension of the epic, making it easier to arouse the emotional resonance of English readers.

The high frequency of "horse" (207 times) is far higher than that of "horse racing" (98 times) in Chinese. This difference is related to the cultural symbol cognition of "horse" in English. In Western culture, "horse" is often associated with "chivalry and hero's journey". By highlighting "horse", the translator transforms the specific event of "horse racing" into a more universally meaningful image of "hero and steed", enhancing cross-cultural resonance. The concentrated appearance of words such as "fight", "war", and "hero" (36-38 times each) reflects English's reliance on "action words". Compared with Chinese, which implies conflicts through narrative scenes, English tends to use direct action words (such as "fight" and "war") to strengthen the dynamic sense of the plot. This adjustment makes the translation more in line with English readers'reading expectations of epics as "fast-paced and explicitly conflictual".

In summary, the high-frequency words in the

original Chinese text reflect "cultural nativeness". By emphasizing core characters, regions, and cultural symbols, it constructs an authentic epic context. In contrast, the high-frequency words in the English translation are adjustments oriented to the target language. Through priority restructuring, image integration, and cultural translation, it realizes the effective dissemination of the epic in the English-speaking world. The essence of the differences between the two lies in the tension between "adherence to the source language culture" and "adaptation to the target language communication".

4.3 Lexical Complexity

4.3.1 Original Chinese text

The average word length of the original Chinese text is 2.3 characters, with a wide distribution of lexical difficulty levels. It includes both simple and commonly used words, as well as relatively complex and obscure ones—especially terms related to Xizang culture, history, and religion, which are more difficult. The use of these complex words enhances the cultural connotation and depth of the text, reflecting the unique charm of King Gesar as a national epic.

4.3.2 English translation

The average word length of the English translation is 5.2 letters, differing from that of the original Chinese text due to the structural differences between English and Chinese. In terms of lexical difficulty, low-difficulty words account for a relatively high proportion in the English translation, while high-difficulty words make up a low proportion. This indicates that during the translation process, the translator took into account the language proficiency and reading habits of target readers, appropriately reducing lexical difficulty and using more accessible words. For some complex culture-loaded terms, the translator adopted methods such as annotations and free translation to lower the reader's comprehension difficulty.

The differences in lexical complexity between the original Chinese text and the English translation are mainly determined by the characteristics of the two languages and the purpose of translation. As a national epic with profound cultural heritage, the original Chinese text requires a certain number of complex words to convey its cultural connotations. The main purpose of the English translation, however, is to disseminate this epic to readers in the English-speaking world. Therefore, appropriate adjustments to lexical difficulty are necessary to improve the acceptability of the translation.

5. Conclusion

This study adopts quantitative linguistic methods to conduct a comparative analysis of the original Chinese version of King Gesar retold by Alai and its English translation by Howard Goldblatt and Sylvia Lin in three dimensions: lexical richness, word frequency, and lexical complexity. The core conclusions are as follows:

The type-token ratio (TTR = 0.11) and standardized type-token ratio (STTR = 0.092) of the original Chinese text are significantly higher than those of the English translation (TTR = 0.07, STTR = 0.05). This indicates that the original Chinese text constructs delicate narrative layers and cultural expressions through richer lexical diversity. In contrast, the English translation tends to use more common words to adapt to the reading habits of English readers, reducing lexical diversity to a certain extent.
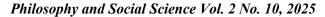
The original Chinese text has an average word length of 2.3 characters, including a large number of Xizang cultural terms with a wide distribution of lexical difficulty. The English translation has an average word length of 5.2 letters, but low-difficulty words account for a higher proportion. It simplifies culture-loaded words through free translation, annotations, and other methods, reducing the overall lexical difficulty to better match the cognitive level of English readers.

In summary, on the basis of retaining the core narrative and spiritual essence of the original work, the English translation achieves a balance in cross-cultural communication through "adaptive adjustments" at the lexical level. It not only maintains the epic's heroic theme and mythological color but also improves the acceptability of target language readers through linguistic simplification and cultural translation.

6. Research Limitations and Future Prospects

This study only focuses on quantitative analysis at the lexical level, without involving comparisons at higher levels such as sentence structure and textual organization. The corpus selected is Alai's retold version, and no comparison is made with translations of traditional oral versions.

Future research can combine textual linguistics to analyze the narrative structure differences

between the original work and the translation, compare the evolution of lexical strategies in different translations, investigate English readers' understanding of cultural words, and verify the actual communication effect of translation strategies.

References

[1] Li, Q. (2011). A review of English translation studies on Li Qingzhao's ci-poems. In Proceedings of the 1st Tsinghua Asia-Pacific Forum on Translation and Intercultural Studies (pp. 84-85). School of Foreign Languages, Zhejiang Gongshang University.

[2] Qin, H. W., & Wang, K. F. (2012). A corpus-based study on the lexical features of translated texts. Chinese Translators Journal, (1), 16-23.

[3] Zhang, M. F. (2015). A study on Howard Goldblatt's translation style: A case study of the English translation of King Gesar. Foreign Language Teaching Theory and Practice, (2), 78-84.

[4] Yang, S. H. & Yang, Y. Q. (2016). Exploring a "shortcut" to achieving faithfulness in linguistic style: A case study of the English translations of the Ming-Qing novel "Sanyan". Chinese Translators Journal, 37(3), 101-105.

[5] Liu, H. T.& Fang, Y. (2017). The principle of faithfulness and stylistic variation in poetry translation: A case study of three English translations of Stray Birds. Journal of Zhejiang University (Humanities and Social Sciences Edition), 47(4), 89-103.

[6] Zhang, X. J. (2021). Is "Hua'er" from China still "Hua'er" in the English-speaking world? Progress in Quantitative Linguistics, 128-137.

[7] Yang, F. (2024). A preliminary discussion on the English translation of the lyrics in King Gesar Epic with the intervention of English ballads. Studies in Language and Culture, 32(6), 180-185.

[8] Yuan Li & Xiaojin Zhang. (2024). A Quantitative Study on the Linguistic Features of the Creation Epic Hnewo Teyy. Economic Society and Humanities, 1(10).

[9] YinnanXiang. (2024). A Quantitative Study on the Linguistic Features of the Tujia Epic Sheba Songs. Advances in Humanities Research, 10(1), 43-52.

[10] Danzhu Angben. (2025). Grasp the foundation, gain a profound understanding, and inherit and promote the epic Gesar. Journal of Qinghai Normal University (Social Sciences Edition), 47(1), 1-13+173.