

Research Article

Corpus-Based Research Methodologies in Translation Studies: Issues, Applications, and Prospects

Xiaqing Li¹ , Huachun Chen¹ , Zhifeng Kang^{2, 3, *} 

¹School of Foreign Languages, Shanghai University, Shanghai, China

²School of AI Translation, Interpreting and Communication, Heilongjiang International Studies University, Harbin, China

³College of Foreign Languages and Literature, Fudan University, Shanghai, China

Abstract

The adoption of corpus-based research methodologies in translation studies has prevailed persistently, from which substantial findings have been drawn both internationally and in China. This paper examines this trend by reviewing representative corpus-based research in translation studies, with particular focus on translator style touching translating Chinese classics and on Chinese-foreign language contexts for Chinese and international publications respectively. The findings indicate that the methodologies have incorporated a quantitative perspective, featuring statistical presentation of data, into traditional qualitative analysis, resultantly enhancing the objectivity and credibility of pertinent research. Despite the advantage, certain limitations remain, including restricted multimodal and multilingual capacity, the lack of data consideration and large-scale standardized bilingual corpora, and an overemphasis on quantification, etc. Therefore, this paper contributes by underscoring corpus-based methodologies as reliable and versatile tools able to strengthen translation studies and highlighting the benefits of interdisciplinary innovation and competitiveness in digital humanities for future relevant research.

Keywords

Corpus-based Research Methodologies, Translation of Chinese Classics, Translator Style

1. Introduction

The application of corpus-based research methodologies in translation studies formally began with the publishing of Mona Baker's seminal article *Corpus Linguistics and Translation Studies: Implications and Applications* at the University of Manchester in 1993 [14]. The methodologies are applied under the guidance of linguistic and translation theories, in which source and target texts are treated as primary research data and analyzed through statistical means. Specifi-

cally, they serve for collecting and processing textual data, and providing corresponding statistical descriptions, thereby facilitating the diachronic or synchronic analysis of translational activities and the elucidation of rationales behind translation phenomena, through which systematic inquiry into the nature of translation is made possible [7, 9]. Thus came the commencement of corpus-based translation studies, and over the past three decades, this research field has seen abundant

*Corresponding author: kangzhifeng@fudan.edu.cn (Zhifeng Kang)

Received: 21 August 2025; **Accepted:** 9 September 2025; **Published:** 30 September 2025



Copyright: © The Author(s), 2025. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

academic outcomes worldwide. Notably, Mona Baker and her team have conducted much investigation of the commonality of translation and translator style via corpora. They propose that a translator's style can be represented through individual linguistic features in their translations, and conclude that habitual, subconscious language use often distinguishes one translator from another [7, 12, 13]. In recent years, the application has even gained momentum as corpus-based research in translation studies continue to be published on major international journals.

In China, with the continuous introduction of the methodologies, scholars in translation studies are able to approach translation via the combination of both quantitative and qualitative methods in lieu of the single qualitative perspective, which has improved the objectivity of translation studies in China and prompted this discipline to progress beyond the subjectivity and arbitrariness of its traditional qualitative studies [14]. Consequently, more in-depth translation research has emerged, especially with attention to translator style, translation-specific linguistic features, and translation strategies and techniques, ushering in a new developmental phase for Chinese translation studies [7]. Furthermore, against the broader backdrop of globalization and China's entering the New Era, it is becoming exceedingly crucial to conduct translation studies with Chinese characteristics to develop a discourse system that bridges China and the world [4]. Much effect has been made, among which the translation of Chinese classics, as an essential channel for promoting Chinese culture abroad, has received heightened scholarly interest. By employing corpus-based methodologies for studying Chinese classics translations, fruitful results have been yielded from research on the style of different expert translators at rendering Chinese classics [2, 6].

Therefore, this paper selects six articles in which the methodologies are adopted for examination: three from CSSCI-indexed journals that focus on translator style in translating Chinese classics, along with another three involving Chinese-foreign language contexts and published internationally on a SSCI-indexed journal. It then delivers reviews of how corpus-based methods are utilized in details, followed by result sections presenting their findings and analyses, and lastly offers evaluations and implications in Discussion.

2. Methodology

The Chinese publications were retrieved from CNKI, the most used and authoritative database in China, whereas the international ones were obtained from Scopus, a world-renowned and much-trusted source of journal articles. Keywords for search included "corpus-based methodologies" "translation studies" "translation of Chinese classics" etc. and their Chinese counterparts, with the results sorted by relevance, newest dates, and CSSCI or SSCI-indexed journals only. Then six of them were chosen which are expected to

reflect the latest trend in recent 10 years of the adoption of corpus-based methodologies in Chinese translator style studies on Chinese classics translation and in international translation studies concerning Chinese-foreign language contexts.

3. Literature Review

3.1. Corpus-Based Methodologies in Translator Style of Chinese Classics Translation

Chinese classics, as the national treasure of Chinese literature and history, have appealed to numerous translators at home and abroad, who have produced plenteous translations of the classics. Predicated on these translations as research data, studies on the style of Chinese classics translators tend to adopt corpus-based methodologies for their capability of processing multiple full versions and analyzing the intrinsic qualities of language on various textual levels. Following is a review of three representative studies of this research area which have conducted solid analyses via corpus-based methods and offered reliable findings and implications.

3.1.1. Corpus-Based Studies of Translator Style in the Translation of Chinese Classics

Among the vast sea of Chinese classics, *the Analects* and *Tao Te Ching*, which encapsulate the essence of the two most prominent Chinese beliefs respectively, Confucianism and Taoism, have been favored by both translators and scholars. Zhao [13] investigated translator style taking Wu Jingxiong's and Arthur Waley's English translations of *Tao Te Ching* as research subjects. In addition to constructing parallel corpora of the two translations, a reference corpus of 2300 commonly used English words was also applied in order to examine their differences with respect to "simplification" and "clarification." WordSmith was selected as the tool for processing the corpus data, and its quantitative results would present differences between the two translations regarding vocabulary, total number of sentences, mean sentence length, frequency of high-frequency words, and personal pronoun usage, all of which were then further supplemented by a qualitative textual analysis that illustrated specific translation choices. Lü and Chen [7] pivoted their research around Gu Hongming's and Arthur Waley's English translations of *the Analects*. Following the building of parallel corpora of the two translations was a quantitative analysis of the text data on the lexical, sentential, and discourse levels, done with the assistance of the corpus retrieval and analysis tools WordSmith and AntConc. Their analysis relied heavily on several parameters: The ones on the lexical level included standardized type-token ratio (STTR), lexical density, and mean word length. For comparison, the TTR (type-token ratio) reflects a text's lexical richness, while STTR, representing the ratio of types to tokens per thousand words in a text, rules out the influence of word count discrepancies between texts on data outcomes.

Lexical density measures the proportion of content words, indicating both the amount of information and the relative difficulty of the text, whereas mean word length indicates stylistic tendencies in wording. On the sentence level, sentence count, mean sentence length, and sentence standard deviation were the common indicators, and concurrently on the discourse level, the indicators comprised conjunction usage and cohesion. Zhang et al. [12] conducted a comparative study using four English versions of *Tao Te Ching* translations. Translations by James Legge, Arthur Waley, Gu Zhengkun, and Xu Yuanchong were selected as the research objects, the anatomy of which depended on the similar research framework and analytical tools used in Lü and Chen's study. The significance of this study lay in its choice of representative translators from both China and the West across different historical periods, and thereby the use of corpus-based methods was applicable to quantifying both the overall stylistic evolution of *Tao Te Ching* translations over time, and the diachronic contrasts between Chinese and Western translators.

3.1.2. Results

Zhao's results [13] are as follows. At the lexical level, Wu's translation exhibited a TTR of 21.54, higher than Waley's 19.03. When compared with the reference corpus, Wu's version was found to use more common vocabulary, indicating that its lexical richness lay particularly in common and functional words. The author then further cited both translators' renderings of the line “其政闷闷，其民淳淳；其政察察，其民缺缺” (Wu's version: If a ruler is mum, mum;//The people are simple, simple.// If a ruler is sharp, sharp, //The people are wily, wily. Waley's version: When the ruler looks depressed, the people will be happy and satisfied;// When the ruler looks lively and self-assured, the people will be carping and discontented), and pointed out that Wu's translation used words with fewer letters, thereby producing a version that could be considered clearer and simpler. At the sentence level, Wu's translation contained 750 sentences with a mean sentence length of 11.25 words, while Waley's included 528 sentences averaging 17.66 words, suggesting that Waley's syntactic structures were more complex. Furthermore, the conjunction “that” appeared 164 times in Waley's version, far exceeding Wu's 40 occurrences. By exemplifying Waley's use of clauses starting with “that” as attributive components for translation, the author argued that Waley's language could be characterized by greater rigor and a more formal register. At the discourse level, Wu's translation contained 78 occurrences of “you” and 75 of “he”, whereas Waley's had 31 occurrences of “you” and 135 of “he”. Together with illustrative examples of differences in pronoun use, these findings suggested that Waley created an objective, declarative narrative context, and by contrast Wu fostered a communicative and dialogic context. For these stylistic differences, the author interpreted them through the translators' perceptions of *Tao Te Ching*, as well as their translation purposes and strategies.

The explanation was that, Wu, who believed that Taoist culture and Christian culture shared the same origin, attempted to get across his perception to the West. Resultantly, he employed the method of free translation to enhance the acceptability of his translation. Waley, in contrast, as a Sinologist who categorized *Tao Te Ching* as a cultural-historical text, placed greater emphasis on content than on form, and therefore favored literal translation in pursuit of maximum fidelity to the original.

Results from the corpus analysis in Lü and Chen's research [7] showed that Gu's translation scored higher than Waley's in terms of STTR, lexical density, mean sentence length, sentence standard deviation, and conjunction ratio at 59.16, 0.713, 17.74, 11.08, and 2.19 respectively, compared to Waley's at 58.68, 0.702, 16.43, 10.05, and 2.04. Waley's translation, however, recorded higher values in mean word length and total sentence count at 4.35 and 1,681, versus Gu's at 4.13 and 1,652. What was implied by the results demonstrated that Gu's version used more varied vocabularies, conveyed more information, and employed more complex sentence structures; meanwhile, Waley's version adhered more closely to the source text and favored shorter sentences. Subsequently, the researchers concluded that Gu's style was characterized by liberal, flexible, and explicative tendencies, whereas Waley's style leaned toward phonetic rendering, literalness, and concision. The explanation of this difference, after the examination of how the two translators rendered the culture-loaded terms, was extracted from the perspective of translation strategy. The researchers argued that Gu, well-versed in both Chinese and Western traditions, adopted domestication in fluent English to render *the Analects* more accessible to Western readers. On the contrary, Waley, who aimed to preserve the linguistic characteristics of the Chinese original, adopted the foreignization strategy more frequently.

The study by Zhang et al. [12], following the same research patterns of Lü and Chen's, concluded from the results that at both the lexical and syntactic levels, the translation styles of *Tao Te Ching* renderings trended toward concision and toward targeting general readerships, which had been conducive to the international dissemination of the classical work. Particularly, the versions produced by Chinese translators were found to be more concise in vocabulary and syntax than those produced by Western translators. To elucidate these diachronic tendencies and the divergences between Chinese and Western versions, a qualitative analysis of the translations in relation to the translators' historical contexts and translation purposes was necessary. The researchers pointed out that the Chinese translators, motivated by the goal of rendering Chinese culture more readily acceptable to Western readers, set their translations of *Tao Te Ching* to be increasingly concise and lucid. By contrast, Western translators, who were under the influence of Christianity and observed the heightened Western interest in China in the post-WWII era, tended to preserve or even amplify Chinese cultural elements in their translations. This practice, while enriching the cultural di-

mension of the translations, could somewhat diminish the concision in style.

3.2. Corpus-Based Methodologies in International Translation Studies

In International publications concerning translations studies, it has long been established to include corpus-based methods as methodologies, with the assistance of which translation as an intricate cross-cultural phenomenon has been further explored and explained. Below is presented a review of three solid studies where corpus-based methods, playing a prominent part in more sophisticated quantitative analyses of a higher number of texts, are combined with a larger variety of theoretical frameworks.

3.2.1. Corpus-Based Studies on Translation Strategies and Ideological Implications in Chinese-Foreign Language Contexts

Casas-Tost's research [1] features an investigation into the translation of Chinese onomatopoeia into Spanish by means of corpus-based methods. For corpus construction, the study compiled data from seven Spanish translations of contemporary Chinese novels. The selected works were authored by different Chinese writers, and translated by different Spanish translators, all of whom adhered to the Chinese originals. These criteria expectedly maximized the representativeness of the sample. The study was grounded in Toury's adequacy-acceptability conceptual framework and Molina's taxonomy of 18 translation techniques, demonstrating a compound use of theories. Zhang et al. [11] applied the methodologies to analyze the use of personal pronouns in children's literature translated from English into Chinese, in the hope of exploring the phenomenon of explicitation in the translations. The study built the following parallel corpora: the Translated Chinese Children's Literature Corpus (TCCLC), and the Non-translated Chinese Children's Literature Corpus (NCCLC). The former comprised 22 English-to-Chinese translated children's books, representing the body of translated Chinese children's literature and totaling 1,168,137 tokens, and concurrently the latter consisted of 20 original Chinese children's books, which symbolized native Chinese children's literature and amounted to 1,215,259 tokens. The target readership of these works was uniformly children aged 7 to 11. For preliminary data processing, all PDF files were converted into TXT format, manually proofread, and segmented via SegmentAnt. For corpus analysis, WordSmith served for the calculation of the frequencies and distribution of personal pronouns in both the TCCLC and NCCLC. On this basis, an independent-sample t-test was also conducted to assess whether the differences between the two corpora were statistically significant, which showcases an intersection of statistics and translatology. Li and Pan's study [3] involves the corpus-based approach to examine the construction of China's image by analyzing the English translations of Chi-

nese political discourse. Their research adopted Systemic Functional Linguistics' Appraisal Theory and van Dijk's Ideological Square as its theoretical frameworks. It established a bilingual parallel corpus with work reports and white papers from the Chinese authorities since 2000 and their corresponding English translations. WordSmith, Excel, and Para-Conc were taken as research tools to organize and extract corpus information, which ultimately yielded 334 appraisal epithets together with their corresponding contexts. These epithets were then classified into three subsystems: attitude, graduation, and engagement, with which the researchers subsequently conducted a qualitative analysis to determine the translation strategies employed by the translators for different categories of appraisal epithets.

3.2.2. Results

Analysis in Casas-Tost's study [1] revealed that only 16.7% of Chinese onomatopoeic expressions were rendered as Spanish ones, with 32.6% of them omitted, and the remaining 50.6% were translated using other lexical categories in Spanish. The researcher elaborated these findings with examples illustrating the specific circumstances and strategies of onomatopoeia translations. In relation to the findings, the study drew on Toury's framework, interpreting the results in terms of translators' individual decision-making rather than systemic differences between the two languages. Both statistical and qualitative analyses were conducted for each of the seven translators. The results further showed considerable variation across the translators: Fisac, Alonso, and Preciado y Hu leaned more toward adequacy, employing a range of strategies to render Chinese onomatopoeia; conversely, Espín and Eherenhaus inclined more toward acceptability, opting to translate Chinese onomatopoeia primarily with expressive words in Spanish rather than direct onomatopoeic equivalents; Ramírez and Suárez occupied intermediate positions on the continuum from adequacy to acceptability. The study thus arrived at the conclusion that the translation of Chinese onomatopoeia into Spanish was influenced less by discrepancies in the size of the onomatopoeic lexicons of the two languages than by translators' strategic choices. It further identified the methods most frequently employed by translators when rendering onomatopoeic expressions as reduction and substitution.

The test results given by Zhang et al. [11] showed $t = 3.65$, $p < 0.001$, confirming that the difference in personal pronoun use between the two corpora was statistically significant. More to the point, the mean standardized frequency of personal pronouns was 83.89 per thousand words in the TCCLC, compared to 62.47 per thousand words in the NCCLC. This indicated that children's literature translated from English into Chinese exhibited a marked explicitation of personal pronouns compared to original Chinese children's literature. To account for this, the researchers carried out qualitative analysis in which previous studies were referenced. They argued that the explicitation of personal pronouns in translated children's literature could be attributed to translators' considera-

tion of young readers' comprehension abilities, as well as the subconscious influence of the high frequency of personal pronouns in English. Furthermore, statistical findings exhibited that the inclusive first-person plural pronoun "zánmen" (咱们) did not appear at all in the TCCLC, whereas in the NCCLC it occurred at a frequency of 0.1 per thousand words. Through qualitative interpretation, the researchers explained this disparity from the perspective of the Unique Item Hypothesis, a theory positing that target-language-specific items lacking a direct equivalent in the source language are often underrepresented in translations. In Chinese, "zánmen" conveys the inclusion of readers, listeners, or interlocutors, an additional nuance not expressed by "wǒmen" (我们, "we"), which the English "we" also lacks. Consequently, the unique item "zánmen" was entirely absent from the TCCLC.

Through the process, Li and Pan [3] discovered that equivalent translation was the most frequently used strategy across all subsystems and their subcategories; zero translation was predominantly used within the "contraction" subcategory

of the engagement subsystem and for positive epithets in the graduation subsystem; moreover, shifting translation was also applied to epithets in the "contraction" subcategory of the engagement subsystem and the "force" subcategory of the graduation subsystem. Hinging on the theory of the Ideological Square and a detailed analysis of examples, the study concluded that these methods reflect translators' efforts to remain largely faithful to the source texts, especially given the texts' high authority. However, the influence of a translator's personal ideology should not be underestimated. The use of zero translation, which was not uncommon, was attributed to some translators who felt they should adhere to English linguistic conventions and avoid excessive use of modifiers. This approach, as was argued by the study, did not effectively present a positive image of China in the target language, but rather diluted the Chinese government's achievements and efforts in national governance.

A table synthesizing the six articles is as follows:

Table 1. Summary of the six articles.

Studies	Research focus	Corpus Type	Corpus Size	Analytical Tools	Theoretical Framework	Main Findings and Interpretation
Zhao [13]	Translator style in Wu's and Waley's translations of <i>Tao Te Ching</i>	Chinese and English Parallel Corpora and English Reference Corpus	<i>Tao Te Ching</i> and two translations; 2300 common English words	WordSmith	Descriptive translation studies	Wu's style: simpler, dialogic; Waley's style: formal, objective, literal. Caused by cultural beliefs and translation purposes.
Lv & Chen [7]	Translator style in Gu's and Waley's translations of <i>The Analects</i>	Chinese and English Parallel Corpora	<i>The Analects</i> and two translations	WordSmith, AntConc	Descriptive translation studies	Gu's style: liberal, explicative; Waley's style: concise, literal. Reflects choice difference of translation strategies.
Zhang et al. [12]	Diachronic and cross-cultural translator styles in <i>Tao Te Ching</i> translations	Chinese and English Parallel Corpora	<i>Tao Te Ching</i> and four translations	WordSmith, AntConc	Descriptive translation studies	Overall trend toward concision; Chinese translators' style more concise; Caused by influence of different cultural and historical contexts.
Casas-Tost [1]	Translation of Chinese onomatopoeia into Spanish	Chinese and Spanish Parallel Corpora	Seven contemporary Chinese novels and their Spanish translations	Manual operation (owing to the lack of analytical tools for the Chinese-Spanish language pair)	Toury's adequacy-acceptability; Molina's translation techniques	Only 16.7% rendered as Spanish onomatopoeia; translators vary on the adequacy-acceptability continuum for translation choice.
Zhang et al. [11]	Explication of personal pronouns in children's literature translated from English into Chinese	Translated Corpus (TCCLC) and Chinese Comparable Corpus (NCCLC)	Twenty-two English-to-Chinese children's books and twenty original Chinese children's books	WordSmith, SegmentAnt	T-test; Unique Item Hypothesis	Significant explication of pronouns in translations; absence of "zánmen" explained by Unique Item Hypothesis.
Li & Pan	Image building in	Chinese and English	Work reports	WordSmith, Ex-	Appraisal Theo-	Equivalent translation most

Studies	Research focus	Corpus Type	Corpus Size	Analytical Tools	Theoretical Framework	Main Findings and Interpretation
[3]	English translations of Chinese political discourse	Parallel Corpus	and white papers since 2000 and their translations	cel, ParaConc	ry; Ideological Square	commonly used, followed by zero translation and shifting; zero translation damaged image building.

4. Discussion

4.1. Corpus-Based Methodologies in Translator Style Studies of Chinese Classics: Reliability, Validity, and Challenges

It is stated that, traditionally, studies of translator style were predominantly conducted by qualitative methods, which examined the overall characteristics of translations from outside the language system [7]. However, the limitation of pure qualitative approaches resides in their inability to provide objective and detailed descriptions of stylistic features grounded in the language itself. Conversely, corpus-based methods enable statistical analysis of translations based on linguistic data and permit interpretation of the data from within the language system. This, to some extent, offsets the subjectivity and partiality of traditional qualitative studies, thereby rendering research findings more objective and scientific [7]. In other words, corpus-based methodologies have not only enriched the research tools and perspectives available for studying translator style, but also reoriented similar studies toward the textual-linguistic level of translations, and thus the conclusions extracted from the studies can take on more persuasiveness.

Reliability and validity are the two crucial criteria for carrying out research. In terms of reliability, the application of corpus-based methodologies in translator style studies has already established a standardized set of procedures. Researchers are generally required to convert translations into TXT files, process and align them using tools such as AB-BYY Aligner in combination with manual adjustment to construct corpora, and then adopt corpus retrieval and statistical analysis software to conduct multi-level quantitative analyses. Typically, the quantitative dimensions include vocabulary (e.g., standardized type-token ratio, lexical density, word length), sentences (e.g., mean sentence length, sentence count, sentence standard deviation), and discourse (e.g., cohesion, pronoun usage). Finally, these parameters are supplemented by qualitative analyses to explore the rationales of observed differences. Such research paradigms, continuously applied across different research, have successfully yielded reliable analyses of translator style, which demonstrates that corpus-based methods are both consistent and highly repli-

cable, and therefore reliable. As for validity, corpus analysis tools such as WordSmith and AntConc, having undergone years of development, are now maturely applied, capable of accurately computing and presenting quantitative textual information. Thus, corpus-based methodologies possess strong validity.

Nevertheless, the application of corpus-based methodologies in translator style studies concerning translating Chinese classics faces several pressing challenges [10]. First, most corpora of classical text translations are of small scales and constructed by scattered researchers, the academic field still lacking large-scale, standardized Chinese-English parallel corpora. This limitation risks confining research to a limited number of translators, a specific time period, or one version of translations while overlooking other translators' works or multiple versions by the same translator, thereby reducing representativeness and increasing the danger of homogenization in research outcomes. Second, current corpus methodologies remain restrained by a unimodal form, with input formats restricted to text documents, unable to incorporate or process multimodal data such as images, color, or sound. This restriction denotes that translator style studies may neglect the rendering of information beyond written text, narrowing the scope of Chinese classics translation studies down to culturally oriented textual works and making it difficult to extend corpus-based methods to traditional Chinese science classics or other multimodal texts. Third, the dominance of corpus methodologies has fostered a tendency toward "quantification over qualification," where researchers place excessive emphasis on quantitative analysis while reducing or even overlooking qualitative interpretation. Fourth, as stated by Liu and Huang [5] that investigation into the translator styles of much-acclaimed sinologist doubling as translators can contribute significantly to the acceptability and dissemination of Chinese literature in the Western literary sphere, similar investigation that explores the styles of renowned Western translators is supposed to be done in Chinese classics translation studies; however, the existing studies have not yet attached sufficient importance of such investigation, nor have corpus-based methods been utilized enough in analyzing more specific textual levels in detail, for instance, conjunctions, when Chinese classics translations by Western translators are being studied.

Therefore, scholars seeking to enhance their studies of translator style in Chinese classics translation studies through corpus-based methods should strengthen their digital human-

ities competence, and engage themselves more in interdisciplinary collaboration with research fields such as digital information technology. They should also acquire deeper expertise in the use and development of corpus-related technologies and software toolkits, even if they are more frequently used in other disciplines, so as to introduce larger and more diverse datasets to their own use and heighten the reliability and validity of their research [8]. Furthermore, building collaborative research efforts is essential to broaden the scope of studies across more translators and translation works, so that more comprehensive descriptions and deeper exploration of translation patterns can be enabled, by which studies on translator style in Chinese classics translation can further arise, take shape, and manifest.

4.2. Rethinking Corpus-Based Translation Studies: Language Coverage and Translator Backgrounds

The above reviewed international papers generally exhibit a significant crossing-over-discipline style, which aligns with the current interdisciplinary trend in translation studies and enriches research perspectives. The languages studied are also diverse, not limited to English, which expands the scope of research in the field. The application of the corpus-based methodologies in these studies also demonstrates high reliability and validity. In particular, the study by Zhang et al. [11] used an independent samples t-test to enhance the method's validity, providing an excellent reference for other translation studies that employ corpus-based methods. However, two issues deserve attention from the researchers. First, the language pairs well-supported by corpus data retrieval and analysis software remain limited, mainly confined to those used in developed countries. This likely results in corpus-based translation studies' focusing primarily on mainstream language texts, and translational texts and phenomena of other languages are thus neglected, left with the limited manual corpus processing methods and traditional qualitative analysis. This invisibly exacerbates the inequality of languages and discourse power both in the real world and in the academic field of translation studies. Second, in terms of selecting source texts for corpus data and describing the data, there is still space for more consideration for the translators' backgrounds. For example, the study by Zhang et al. [11], when selecting English-to-Chinese children's literature as corpus data, seems to omit information about the Chinese translators' backgrounds. This led to a probable flaw in the explanation based on Unique Item Hypothesis of the phenomenon that “zánmen”, an inclusive first-person plural form, did not appear in the TCCLC but did appear in the NCCLC. This is because native Chinese speakers who grow up in southern China rarely use the pronoun “zánmen”; strictly speaking, within the Chinese language system of southern China, “zánmen” is not a unique item. If the translators of the translations processed in TCCLC are raised or live in southern

China, it would be natural for them not to use “zánmen” to translate “we”.

To tackle the limitations, it is of urgent necessity to develop corpus retrieval and analysis software that is compatible with more languages, so as to allow researchers of other languages in translation studies the opportunity of using corpus methods to study translation in their respective languages. Simultaneously, it is highly recommended that for data collection, the researchers consider the data to be included in the corpus more carefully and comprehensively, especially with respect to the translators' backgrounds, to ensure the data is more representative and the qualitative analysis is better completed.

5. Conclusion

This paper reviewed several publications predicated on corpus-based methods: three articles which study translator style of Chinese classical text translation and were published in CSSCI-indexed journals, as well as three articles published internationally that investigate Chinese-foreign language translation-related subjects. It reveals that the adoption of corpus-based methodologies in translation studies has formed an applicable and trustworthy set of norms: There are tested procedures and shared software for data input, corpus construction, as well as corpus data processing and analysis, which demonstrates a high degree of reliability and validity. The adoption enables researchers of the field to conduct large-scale collection, organization, and analysis of translated texts, thereby improving the persuasiveness, scientific rigor, and objectivity of their translation studies. Furthermore, through the comparison between the selected Chinese and international publications, it can be summarized that while international publications also utilize corpus-based methodologies, they tend to have more well-elaborated qualitative analysis sections. This is perhaps related to the different layout requirements of Chinese and international journals, and yet the overemphasis on quantitative analysis warrants the attention of scholars publishing in China. For the studies of translator style in the translation of Chinese classics using corpus-based methods, the main problems currently consist in the lack of large-scale, standardized bilingual parallel corpora for classical texts, the limitation of corpus input formats to TXT files, which could lead to the neglect of traditional scientific and technological Chinese classics, and the partial neglect of qualitative analysis. These problems are likely to be detrimental to the further development of related research. For international translation studies that employ the methodologies, the identified problems can be condensed to the limited number of languages supported by corpus retrieval and analysis software, and the lack of comprehensive consideration for translators' backgrounds in the course of collecting translated texts as corpus data. Thereby, this review makes the following suggestions. In the first place, researchers aspiring to adopt corpus methodologies should improve their competence in digital humanities, interdisciplinary literacy, and

research collaboration. It is also advised that they enhance their capabilities in using corpus technologies and tools to expand the scale of corpora. They are further expected to establish a comprehensive method for evaluating translated texts as corpus data and place more emphasis on qualitative analysis. Only in this way can they more comprehensively and accurately describe translational phenomena and investigate the causes, therefore better grasping the underlying laws and essence of translation.

Abbreviations

STTR	Standardized Type-Token Ratio
TTR	Type-Token Ratio
TCCLC	Translated Chinese Children's Literature Corpus
NCCLC	Non-translated Chinese Children's Literature Corpus

Funding

This paper is funded by 2022 Chinese Fund for the Humanities and Social Science (22WZSB029), 2023 "Training and Development Program for Young Teachers in Shanghai Universities" of the Shanghai Municipal Education Commission: An Exploration of Ideological and Political Education in Multimodal Translation Courses during the Transition Period and Fudan University's First Series of "Seven Premium Textbook Projects": Embodied-Cognitive Interpreting Studies.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Casas-Tost, H. (2014). Translating onomatopoeia from Chinese into Spanish: a corpus-based analysis. *Perspectives*, 22(1), 39-55. <https://doi.org/10.1080/0907676X.2012.712144>
- [2] Fan, M. (2017). Ji yu yu liao ku de lun yu wu yi ben wen hua gao pin ci fanyi yan jiu [A study of the translation of cultural high-frequency words in five versions of *the Analects*]. *Waiyu jiaoxue (Foreign Language Education)*, 38(6), 80-83. <https://doi.org/10.16362/j.cnki.cn61-1023/h.2017.06.015>
- [3] Li, T., & Pan, F. (2021). Reshaping China's Image: a corpus-based analysis of the English translation of Chinese political discourse. *Perspectives*, 29(3), 354-370. <https://doi.org/10.1080/0907676X.2020.1727540>
- [4] Li, X. Q., & Zheng, L. (2023). Ji yu VOSviewer de Zhong guo te se fanyi xue ke jian gou shi xiang hua tan ze [A visual exploration of the construction of Chinese-characteristic translation discipline based on VOSviewer]. *Fanyi yanjiu yu jiaoxue (Translation Research and Teaching)*, (2), 113-119.
- [5] Liu, F. R., & Huang, Q. (2024). Ji yu yu liao ku de Bai Ya Ren yi zhe feng ge fen xi yi Yu Hua xiao shuo yi yi ben wei li [A corpus-based analysis of Bai Yiren's translator style: A case study of the English translations of Yu Hua's novels]. *Fanyi yanjiu yu jiaoxue (Translation Research and Teaching)*, (1), 97-107.
- [6] Lou, B. C., & Zhao, D. Y. (2023). Ji yu yu liao ku de meng zi yi yi ben wen ti te zheng du wei fen xi [A multidimensional analysis of the stylistic features of English translations of *Mencius* based on corpus]. *Dangdai waiyu yanjiu (Contemporary Foreign Language Studies)*, (5), 157-166.
- [7] Lv, P. F., & Chen, D. S. (2021). Ji yu yu liao ku de lun yu yi yi ben fanyi feng ge bi jiao yan jiu yi gu hong ming he ya si wei li liang yi ben wei li [A comparative study of the translation styles of English versions of *the Analects* based on a corpus: A comparison of the versions by Gu Hongming and Arthur Waley]. *Shanghai fanyi (Shanghai Journal of Translators)*, (3), 61-65.
- [8] Wang, G. F., & Liu, Y. L. (2022). Yu liao ku fu zhu de fanyi pi pan he fanyi zhi liang ping gu [Corpus-assisted translation criticism and quality assessment]. *Fanyi yanjiu yu jiaoxue (Translation Research and Teaching)*, (2), 99-104.
- [9] Wang, K. F. (2012). Yu liao ku fanyi xue tantao [Explorations in corpus-based translation studies]. Shanghai: Shanghai Jiao Tong University Press.
- [10] Xu, M. W., & Wang, P. (2022). Ji yu yu liao ku de Zhong guo ke ji dian ji yi yi yan jiu xian zhuang yi yi yu wang wang [A corpus-based study of the English translation of Chinese scientific classics: Current situation, significance, and prospects]. *Waiyu yu waiyu jiao xue (Foreign Languages and Teaching)*, (5), 116-124+149. <https://doi.org/10.13458/j.cnki.flatt.004886>
- [11] Zhang, X., Kotze, H., & Fang, J. (2020). Explication in children's literature translated from English to Chinese: A corpus-based study of personal pronouns. *Perspectives*, 28(5), 717-736. <https://doi.org/10.1080/0907676X.2019.1689276>
- [12] Zhang, X. R., Xing, Y. L., Zhang, P., et al. (2022). Dao de jing si ge yi yi ben de fanyi feng ge dui bi yan jiu ji yu yu liao ku de tong ji yu fen xi [A comparative study of the translation styles of four English versions of the *Tao Te Ching* based on a corpus: A statistical and analytical approach]. *Shanghai fanyi (Shanghai Journal of Translators)*, (3), 33-38.
- [13] Zhao, Y. (2015). Ji yu yu liao ku de dao de jing liang yi ben de fanyi feng ge yan jiu [A corpus-based study of the translation styles of two English versions of the *Tao Te Ching*]. *Zhongguo fanyi (Chinese Translators Journal)*, 36(4), 110-113.
- [14] Zhu, Z. W., & Li, R. F. (2023). Ji yu yu liao ku de Pang De Zhong guo dian ji yi yi yi zhe feng ge tan xi [A corpus-based analysis of Pound's translator style in translating Chinese classics]. *Waiyu jiaoxue (Foreign Language Education)*, 44(4), 75-82. <https://doi.org/10.16362/j.cnki.cn61-1023/h.2023.04.018>