Social Sciences and Humanities
ISSN Print: 3104-4328
ISSN Online: 3104-4336
CD ISSN: 3104-4344
https://doi.org/10.63313/SSH.9092

# A Corpus-Based Evaluation of Human and Machine Translation Quality for the Literary Classic *The Analects*

**Yanling Yuan*. Jiaqi Hu and Yimei Xiao**
School of Languages and Literature, University of South China, Hengyang 421001, China
Email: 89718864@qq.com

**How to cite this paper:** Yuan, Y. L., Hu, J. Q., & Xiao, Y. M. (2026). A corpus-based evaluation of human and machine translation quality for the literary classic *The Analects*. *Social Sciences and Humanities*, *4*(1), 48–63. ISSN Print: 3104-4328; ISSN Online: 3104-4336.
https://doi.org/10.63313/SSH.9092
**Published: 2026-05-14**



## Abstract

This study is based on a self-built English translation corpus of *the Analects*, which includes two human translations and two artificial intelligence (AI) translations. By integrating research methods from corpus linguistics and quantitative linguistics, and employing a translation quality assessment model, this study conducts a comparative analysis of the translation quality of human and AI translations from two major dimensions: lexis and syntax. The results show that in the lexical dimension, human translation demonstrates greater flexibility in word selection and lexical richness, and is also more versatile in part-of-speech conversion. In the syntactic dimension, AI translation systems can effectively identify and reproduce the logical structure of the original text, generating logically coherent and structurally complete translations, but they tend to exhibit a certain degree of word-for-word translation. Based on these findings, this study points out that both human and AI translations have their own strengths and limitations. Therefore, this paper proposes strategies to optimize the human-computer collaborative translation model in order to enhance the application efficiency of artificial intelligence in the translation of classical works.

## Keywords

Corpus; Translation Quality Assessment; Quantitative Linguistics; Human-Computer Collaborative Translation

## 1. Introduction

With the deepening implementation of China's "Culture Going Global" strategy, literary classics, as treasures embodying the profound heritage of Chinese civilization, have assumed an increasingly prominent position in domestic translation studies, serving to present Chinese narratives and amplify China's voice to a wider international audience. As a concentrated expression of the essence of Confucian thought, The Analects, with its unique cultural connotations and philosophical insights, stands as an indispensable iconic text in the international dissemination of Chinese culture.

However, given the highly complex nature and profound cultural depth inherent in literary classics, traditional translation software such as Youdao and Google Translate, despite having made certain attempts and efforts in the field of classical translation, still fall short when confronted with the intricate structures and cultural profundity of such texts, often inevitably producing deviations. In contrast, AI translation systems powered by advanced large language models (LLMs), distinguished by their superior comprehensive comprehension, logical optimization capabilities, and cultural adaptability, can offer more precise and efficient solutions for the English translation of classical texts, thereby effectively broadening the channels for the international dissemination of Chinese culture.

In light of the above, this study will draw upon a self-constructed bilingual parallel corpus of The Analects and adopt a methodology integrating corpus linguistics with quantitative linguistics. The research texts selected include the original text of The Analects, D. C. Lau's English translation, Ezra Pound's English translation, the AI translation generated by Wenxin Yiyan, and the AI translation generated by Zhipu Qingyan. First, a comparative analysis of translation quality between human and AI translations will be conducted from the lexical and syntactic dimensions, based on nineteen quantitative linguistic features. Subsequently, a translation quality assessment model will be applied to evaluate the translation quality of human and machine translations and to analyze the underlying reasons for their respective strengths and weaknesses. Finally, corresponding improvement strategies will be proposed based on the advantages and disadvantages identified in both human and AI translations. This study is expected to provide valuable insights for the application of large language models in classical translation and to offer new perspectives and directions for the development of translation in the era of artificial intelligence.

## 2. Research Design

Based on the research subjects, methods, and objectives outlined above, the research design of this study is summarized as follows:

### 2.1. Corpus Selection

In the self-constructed corpus, the original Chinese text selected is The Analects published by People's Literature Publishing House in 2022. The human English translations selected are D. C. Lau's translation (published in 1979) and Ezra Pound's translation (published in 1933). The AI translations selected are the versions generated by Wenxin Yiyan and Zhipu Qingyan. D. C. Lau's translation far outpaces other versions in sales on the Amazon platform and is well received by international readers; Ezra Pound, as a towering figure in the translation field, enjoys high academic recognition for his translations of major classical texts. Wenxin Yiyan and Zhipu Qingyan are cognitive large language models developed by Baidu Online Network Technology Co., Ltd. and Tsinghua University, respectively. Testing has demonstrated their strong comprehension and translation control of classical texts. Therefore, these four translations are selected as the research subjects of this study. This study establishes a one-to-four English-Chinese parallel bilingual corpus comprising the original text of The Analects, two human English translations, and two AI translations, achieving sentence-level alignment between the Chinese original and the four English translations. Both the Chinese original and the English translations have been annotated with part-of-speech tags and segmentation markers.

## 2.2. Selection of Translation Quality Assessment Indicators

Translation quality assessment refers to the process of evaluating translation products according to certain criteria (Dai Guangrong & Zuo Shangjun, 2021: 92). In response to the diversification of assessment models, scholars have proposed that translation quality assessment can be broadly divided into two types: quantitative and non-quantitative models (Dai Guangrong & Zuo Shangjun, 2021: 93). Quantitative models, such as the textual and pragmatic model, apply theories and methods from pragmatics, discourse analysis, cognitive linguistics, and corpus linguistics to translation quality assessment. This study integrates corpus linguistics with quantitative linguistics, adopting a methodology that combines quantitative analysis with textual analysis to assess the quality of the four English translations. A total of nineteen quantitative linguistic features is ultimately identified for comparing the translation quality of the four English versions. At the lexical level, thirteen quantitative linguistic features are employed: standardized type/token ratio, lexical density, proportion of the top thirty high-frequency words in the wordlist, average word length, proportion of nouns, proportion of verbs, proportion of adjectives, proportion of adverbs, proportion of numerals, proportion of measure words, proportion of conjunctions, proportion of pronouns, and proportion of particles. At the syntactic level, six quantitative linguistic features are employed: average sentence length, proportion of declarative sentences, proportion of interrogative sentences, proportion of short sentences (marked by commas and semicolons), and proportion of exclamatory sentences.

## 2.3. Introduction to the Translation Quality Assessment Model

Western scholarly attention to translation quality assessment began in the 1970s, marking a paradigm shift from subjective models of translation criticism to objective models of translation quality assessment (Wu Guangjun, 2007: 75). However, given the diversity of contextual environments, evaluating subjects, and evaluation objectives, different evaluation methods and standards are inevitably adopted in the practice of translation quality assessment, which necessarily results in a multiplicity of discursive forms in translation quality evaluation (Sun Lin, 2023: 37). Consequently, establishing unified and universally applicable standards for translation quality assessment has become particularly complex and challenging.

Drawing upon corpus-based translation studies and quantitative linguistics theories, Qi Yuling and Jiang Yue (2016) collected linguistic quantitative feature data from translated texts using large-scale domestic and international corpora. Through processing and computation, they constructed a reference model for translation quality assessment, thereby achieving quantitative and objective evaluation of translation products. With the aid of software, this model realizes the standardization and automation of the assessment process. Testing has demonstrated that this model possesses certain reference value for the quality assessment of translated texts. In light of this, the present study adopts the aforementioned translation quality assessment model and integrates data from 19 core linguistic quantitative features. Through precise calculation and in-depth analysis, it examines whether the relevant textual features conform to the interval standards stipulated by the model, so as to provide a comprehensive and objective assessment of the translation quality of different human and AI translations of The Analects. Based on data extracted and organized from a substantial body of literature and three large-scale corpora, the numerical intervals for the linguistic quantitative features are obtained as follows:

**Table 1.** Assessment Parameter Intervals for 23 Linguistic Quantitative Features

| Linguistic Quantitative Feature | Assessment Parameter Interval | Linguistic Quantitative Feature | Assessment Parameter Interval |
|---|---|---|---|
| Standardized Type-Token Ratio | 【38.13-55.27】 | Idioms | 【0.53-1.663】 |
| Nouns | 【15.5-22.74】 | Lexical Density | 【41.98-55.59】 |
| Verbs | 【22.93-24.79】 | Word Length | 【1.38-1.74】 |
| Adjectives | 【4.23-6.7】 | Average Sentence Length | 【10.77-29.87】 |
| Pronouns | 【7.0401-11.81】 | Exclamatory Sentences | 【0.1129-0.3115】 |
| Adverbs | 【6.2732-7.5695】 | Periods | 【3.6027-4.1654】 |
| Numerals | 【2.97-3.6279】 | Question Marks | 【0.2916-0.5115】 |
| Measure Verbs | 【2.3186-3.91】 | Commas | 【5.5412-5.8219】 |
| Conjunctions | 【2.61-3.1175】 | Semicolons | 【0.0360-0.1779】 |
| Prepositions | 【4.2445-4.45】 | Enumeration Commas | 【0.0706-0.5573】 |
| Particles | 【7.8706-8.5176】 | Top 30 High-Frequency Words | 36.66 |
| Model Verbs | 【0.5455-1.1876】 | | |

The translation quality assessment model formula is as follows:

**Figure 1.** Assessment Model Formula for Each Parameter

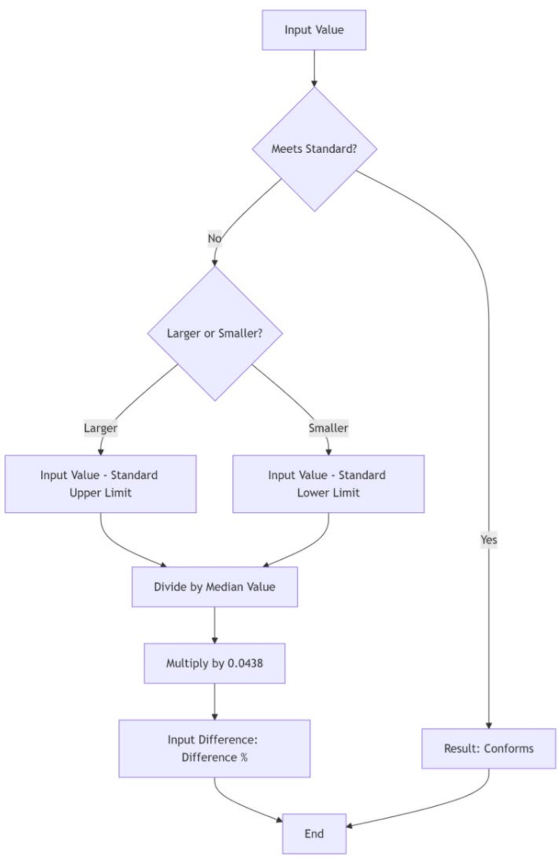
$$score = \sum_{i=1}^{N} C_i (N = 23)$$

among which: $C_i = 0.0438 * \begin{cases} \frac{X_i - S_{(\max)i}}{d_i} (X_i > S_{(max)\mathrm{i}}) \\ 0 \\ \frac{S_{(\min)i} - X_i}{d_i} (X_i < S_{(min)\mathrm{i}}) \end{cases}$

$$d_i = \frac{S_{(\max)i} + S_{(\min)i}}{2}$$

Based on the above assessment formula, and with a view to simplifying the operational process and facilitating the generalizability of this assessment model, Qi Yuling and Jiang Yue (2016) developed a computer program. The underlying logic of the program is to automatically calculate the deviation value for each quantitative feature and the final total deviation value, thereby achieving automation of the computational process. The calculation process for the deviation value of each quantitative feature is illustrated in Figure 2.

**Figure 2.** Calculation Flowchart for Each Parameter




## 2.4. Statistical Tools and Methods

This study employs ABBYY Aligner, Word Smith, CLAWS7, and Em Editor to conduct statistical analyses of the 19 linguistic quantitative features described above. First, Em Editor is used to perform noise reduction (cleaning) on the four translated texts.

ABBYY Aligner is then used to perform sentence-level alignment for each of the four translations. For lexical-level data, the four translated texts are input into Word Smith respectively to obtain the type count, token count, standardized type-token ratio, average word length, average sentence length, and top-30 high-frequency word lists for each translation. Subsequently, CLAWS7 is used to perform part-of-speech tagging on the four translated texts; the quantities of different parts of speech are then counted to derive the proportional distribution of each part of speech across the full text. For syntactic-level linguistic quantitative features, Em Editor is used to count the total occurrences of exclamation marks, periods, question marks, commas, and semicolons in each of the four translations, from which their respective proportional frequencies are derived.

Calculation procedure for the deviation values of linguistic quantitative features: First, the numerical values of each linguistic quantitative feature are compared against the assessment parameter intervals of the model to examine the compliance status of individual parameters. In accordance with the quality assessment model formula above, the deviation values for the 19 linguistic quantitative features are calculated for each of the four translations: the two human translations of The Analects: D. C. Lau's translation and Ezra Pound's translation and the two AI translations: Wenxin Yiyan's AI translation and Zhipu Qingyan's AI translation. Finally, based on the translation quality assessment model formula above, the total deviation value of assessment parameters is calculated for each translation, thereby enabling an objective assessment of the overall translation quality of each version.

## 3. Human-Machine Translation Quality Assessment Based on the Translation Quality Assessment Model

### 3.1. Lexical-Level Investigation

Since variations in lexical features such as type/token ratio (TTR), average word length, and proportions of different parts of speech constitute significant factors contributing to stylistic differences among translations, this section employs Word Smith and Em Editor corpus processing tools to extract lexical-level linguistic features from the D. C. Lau, Ezra Pound, Wenxin Yiyan, and Zhipu Qingyan translations of The Analects. The deviation values are calculated according to the formula presented above, thereby enabling an objective and precise analysis of the differences among the various translations. The specific results are presented in Table 2:

**Table 2.** Assessment Parameter Intervals for 23 Linguistic Quantitative Features

| Linguistic Quantitative Features | D. C. Lau's translation | Assessment Result | Ezra Pound's translation | Assessment Result | Wenxin Yiyan | Assessment Result | Zhipu Qingyan | Assessment Result |
|---|---|---|---|---|---|---|---|---|
| STTR | 37.01 | Compliant | 42.85 | Compliant | 37.86 | Compliant | 35.14 | -0.0028 |
| Nouns | 28.27 | 0.0127 | 30.07 | 0.0168 | 29.97 | 0.0166 | 29.56 | 0.0156 |
| Verbs | 14.46 | -0.0155 | 15.07 | -0.0144 | 16.00 | -0.0127 | 15.94 | -0.0128 |
| Adjectives | 6.21 | Compliant | 6.31 | Compliant | 6.49 | Compliant | 5.94 | Compliant |
| Pronouns | 7.66 | Compliant | 8.73 | Compliant | 9.40 | Compliant | 8.70 | Compliant |
| Adverbs | 4.33 | -0.0123 | 4.21 | -0.0130 | 4.35 | -0.0122 | 3.72 | -0.0161 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Numerals | 0.77 | -0.0292 | 1.04 | -0.0257 | 0.65 | -0.0308 | 0.86 | -0.0280 |
| Conjunctions | 5.45 | 0.0357 | 5.33 | 0.0338 | 7.34 | 0.0645 | 6.82 | 0.0566 |
| Prepositions | 9.89 | 0.0548 | 9.13 | 0.0472 | 8.12 | 0.0370 | 7.40 | 0.0298 |
| Particles | 9.66 | 0.0061 | 8.19 | Compliant | 10.11 | 0.0085 | 10.95 | 0.0130 |
| Model Particles | 0.31 | -0.0117 | 0.10 | -0.0225 | 0.15 | -0.0198 | 0.15 | -0.0202 |
| Word Length | 4.20 | 0.0691 | 4.22 | 0.0696 | 4.49 | 0.0772 | 4.31 | 0.0722 |
| Top 30 High-Frequency Words | 43.56 | 0.0082 | 38.65 | 0.0024 | 41.07 | 0.0053 | 44.09 | 0.0089 |

Based on a meticulous analysis of Table 2, we can clearly observe that the four translations exhibit similar trends across multiple linguistic quantitative features. Specifically, the proportions of adjectives and pronouns remain within the preset model parameter intervals; however, the proportions of nouns, conjunctions, prepositions, particles, word length, and the top 30 high-frequency words slightly exceed the parameter intervals; conversely, the proportions of verbs, adverbs, numerals, and modal particles fall slightly below the preset parameter intervals.

A thorough analysis of this phenomenon reveals its close connection to the distinctive textual characteristics of The Analects. As a collection of sayings compiled by the disciples of Confucius and their successors during the early Warring States period, The Analects primarily records the words and deeds of Confucius and his disciples, concentrating on the political philosophy, ethical morality, and educational concepts of Confucius and the Confucian school. Compared with other literary works, this collection places greater emphasis on the use of nouns for accurate description and conceptual definition, while frequently employing conjunctions, prepositions, and particles to enhance textual coherence and logicality, thereby rendering the text more compact and comprehensible.

### 3.1.1. Lexical Richness

A higher type/token ratio in a text indicates richer vocabulary use and greater lexical variation on the part of the author (Baker 2000: 250). The standardized type-token ratio (STTR) divides the text into segments of a specified length, calculates the type/token ratio for each segment, and finally obtains the average of these ratios. STTR is typically applied to the comparison of texts of different lengths, as variations in the evenness of type distribution across a text directly affect the type-token ratio. Therefore, employing the standardized type-token ratio as an indicator of lexical richness is more reliable and accurate.

As shown in Table 2, with regard to the standardized type-token ratio, both human translations of The Analects and the Wenxin Yiyan translation fall within the model parameter interval, whereas the Zhipu Qingyan translation of The Analects falls slightly below the lower limit of the parameter interval. The specific values are as follows: Ezra Pound’s translation (42.85%) > Wenxin Yiyan’s translation (37.86%) > D. C. Lau’s translation (37.01%) > Zhipu Qingyan’s translation (35.14%). Why does Ezra Pound’s translation exhibit a substantially higher standardized type-token ratio than the other three translations? A close reading of Pound’s translation reveals that the translator has incorporated a substantial amount of annotative material to facilitate better comprehension by target readers. The minimal difference between D. C. Lau's translation and the Wenxin Yiyan translation indicates that human and AI translations may exhibit comparable levels of lexical richness. A comparison between the Wenxin Yiyan and Zhipu Qingyan translations demonstrates that different machine translations also exhibit variations, which may be attributable to the fact that different AI natural language processing systems are based on different large-scale corpora and training data.

#### 3.1.2. Frequency Differences Across Word Classes

To conduct an in-depth investigation into the specific application differences of various word classes in the translations, this study conducts a detailed statistical analysis of the frequency proportions of nouns, verbs, adjectives, adverbs, pronouns, particles, numerals, prepositions, conjunctions, and modal particles. Based on these statistical results, an in-depth analysis of the frequency differences across different word classes is performed, with the aim of comprehensively revealing and examining the distinctive linguistic features exhibited by different translations and further exploring their underlying causes.

With respect to the proportion of noun usage, after controlling for the factor that Ezra Pound's translation contains a large number of annotations resulting in a higher noun proportion, we observe that the noun proportions in the AI translations are higher than those in D. C. Lau's translation. This finding indicates that the AI translations demonstrate relatively strong capability in the processing, recognition, and accurate translation of nouns.

However, in the treatment of verbs, Chinese tends to employ verbs frequently, whereas English tends to favor nominal structures for expression. According to the data in Table 2, the noun proportions in both AI translations are slightly higher than those in the human translations. This reflects that, in order to accommodate the differences between Chinese and English expression, translators in the process of Chinese-English translation not only flexibly convert verbs into nouns but may also convert other word classes into nouns, thereby resulting in relatively higher noun proportions in human translations. This phenomenon also reveals that current AI translation still requires further optimization in terms of word-class conversion functionality to better approximate idiomatic target-language expression habits.

As indicated by the data in Table 2, another area where human and machine translations exhibit substantial differences is in the use of conjunctions. In the comparison between human and machine translations, the difference in conjunction usage is particularly pronounced. Conjunctions, as key formal markers of semantic logical relations, play an indispensable role in constructing coherent organic discourse. This is especially true in English, a language that places equal emphasis on form and function, where the use of conjunctions is of paramount importance. They serve as ligaments that tightly connect sentences and paragraphs, forming texts with clear logic and complete structure. English possesses a greater variety of conjunctions than Chinese, and their frequency of use is considerably higher. The conjunction proportions across the translations are as follows: Ezra Pound's translation (5.33%) < D. C. Lau's translation (5.45%) < Zhipu Qingyan's translation (6.82%) < Wenxin Yiyan's translation (7.34%). It is thus evident that the conjunction proportions in the AI translations of The Analects are markedly higher than those in the human translations, demonstrating that AI translation possesses considerable advantages in ensuring textual logicality and coherence. Through precise identification and employment of conjunctions, AI translation systems can more effectively capture and express the logical relations in the original text, thereby generating more accurate and coherent translations.

## 3.2. Syntactic-Level Investigation

This study statistically analyzes the values of linguistic quantitative features affecting translation quality assessment at the syntactic level such as periods, exclamation marks, question marks, semicolons, and average sentence length across different translations. The resulting data and translation quality assessment outcomes are presented in Table 3:

**Table 3.** Values of Linguistic Quantitative Features at the Syntactic Level and Their Translation Quality Assessment Results

| Linguistic Quantitative Features | D. C. Lau's translation | Assessment Result | Ezra Pound's translation | Assessment Result | Wenxin Yiyan | Assessment Result | Zhipu Qingyan | Assessment Result |
|---|---|---|---|---|---|---|---|---|
| Average Sentence Length | 16.26 | Compliant | 16.46 | Compliant | 14.18 | Compliant | 14.92 | Compliant |
| Exclamatory Sentences | 0.18 | Compliant | 0.08 | -0.0062 | 0.36 | 0.0099 | 0.47 | 0.0323 |
| Periods | 4.97 | 0.0091 | 5.30 | 0.0127 | 5.26 | 0.0123 | 4.76 | 0.0067 |
| Question Marks | 1.05 | 0.0586 | 1.14 | 0.0681 | 1.45 | 0.1026 | 1.51 | 0.1086 |
| Commas | 6.92 | 0.0086 | 6.49 | 0.0052 | 9.53 | 0.0287 | 10.14 | 0.0335 |
| Semicolons | 0.42 | 0.1145 | 1.09 | 0.4071 | 0.47 | 0.1341 | 0.88 | 0.3151 |

### 3.2.1. Average Sentence Length

Upon detailed examination of the data presented in Table 3, although all indicators for the four translations comply with the preset parameter ranges of the translation quality assessment model, human and AI translations exhibit significant differences in the key indicator of average sentence length. Specifically, the average sentence lengths of the two human translations are markedly higher than those of the two AI translations. Scholars such as Hu Kaibao (2016: 13) have explicitly pointed out that a typical characteristic of machine translation is its tendency to treat complete sentences as translation units. Examining the two AI translations, their average sentence lengths are identical, and when closely reading the original text of The Analects, one readily observes that the original sentences are structurally compact and brief in length, a feature commonly found in classical Chinese literary works. This phenomenon indicates that AI translation systems make relatively few modifications to the sentences themselves when processing texts, with limited operations at the sentence level.

In contrast, when comparing D. C. Lau's and Ezra Pound's translations of The Analects, it can be observed that in their pursuit of formal equivalence between the translation and the original text at the sentence level, they tend to increase the structural capacity and average sentence length to assist target-language readers in more accurately comprehending the original content. This practice reflects, to a certain extent, the flexibility and depth of human translation in processing sentence structure and meaning.

In summary, due to the insufficient capability of AI translation systems in the identification and judgment of sentence semantics, they exhibit certain limitations in processing and analyzing sentence structure and meaning, and their ability to deconstruct and analyze sentences remains to be improved. Consequently, the sentence count in AI translations more closely approximates that of the original text, yet has not yet achieved the processing effects of human translation at the sentence level.

#### 3.2.2. Sentence Characteristics

The translation of different sentence types also occupies an important position in the assessment of translation quality. Therefore, this study conducts a detailed statistical analysis of the proportional distribution of periods, exclamation marks, question marks, semicolons, and commas in the translations. Based on the data analysis in Table 3, we find significant differences between human and AI translations in the use of exclamation marks and commas.

In the use of exclamation marks, D. C. Lau's and Ezra Pound's translations are relatively restrained, whereas the machine translations exhibit markedly higher usage rates. Although the original text of The Analects does not contain a large number of exclamatory sentences, these sentences serve as important means for Confucius and his disciples to express strong emotions or profound insights, playing a crucial role in conveying the authors' emotions, intensifying tone, and highlighting themes. They not only reflect the profound reflections of Confucius and his disciples on morality, life, and governance but also demonstrate their distinctive linguistic style and intellectual charisma. Therefore, accurate handling of exclamatory sentence translation is essential for the dissemination of Chinese classical culture. In translation practice, D. C. Lau and Ezra Pound tend to avoid direct correspondence in translating exclamation marks, instead employing more flexible expressive means such as emphatic sentence patterns in English or the addition of intensifying words. In contrast, AI translation often directly replicates the punctuation marks of the original text, which once again exposes the deficiency of machines in comprehensive sentence processing capabilities, with their translation approach often limited to sentence-by-sentence translation. Thus, AI translation systems still require strengthening in the identification and processing of semantic and emotional content in the original text.

With regard to comma usage, AI translations employ commas at a frequency approximately 4% higher than human translations. To analyze the underlying reasons for this difference, we may proceed from the characteristics of The Analects as a classical Chinese text. The grammatical structure of The Analects differs significantly from modern Chinese and English; its sentence structure, though concise, is rich in connotation. When translating it into English, additional commas may be required to supplement or explain certain sentence components in accordance with the grammatical rules and reading habits of the target language, thereby facilitating readers' comprehension and grasp of the original meaning. When processing such complex classical Chinese sentence patterns, AI translation systems tend to employ more commas to separate sentence components in order to ensure grammatical correctness and avoid structural confusion. However, this processing approach may sometimes lead to excessive comma usage. In contrast, when human translators undertake translation, they first develop a thorough understanding of the connotative meaning of the original text, and then express it in more appropriate and natural language. Throughout this process, translators render the translation in accordance with English expression habits and grammatical rules while maximally preserving the characteristics of the original text, thereby ensuring the fluency and accuracy of the translation. In conclusion, although AI translation has demonstrated its efficiency and accuracy in numerous domains, there remains room for optimization when translating literary classics such as The Analects. To enhance the accuracy of AI translation in processing such texts, we need to further optimize translation algorithms and training data, and strengthen the understanding and mastery of background knowledge in classical Chinese and philosophical thought.

## 3.3. Comprehensive Investigation

By integrating the 19 linguistic quantitative features at the lexical and syntactic levels and calculating the total deviation value for each translation, the following results are obtained, as shown in Table 4:

**Table 4.** Values of the 19 Quantitative Features and Their Translation Quality Assessment Results

| Linguistic Quantitative Features | D. C. Lau's translation | Assessment Result | Ezra Pound's translation | Assessment Result | Wenxin Yiyan | Assessment Result | Zhipu Qingyan | Assessment Result |
|---|---|---|---|---|---|---|---|---|
| STTR | 37.01 | Compliant | 42.85 | Compliant | 37.86 | Compliant | 35.14 | -0.0028 |
| Nouns | 28.27 | 0.0127 | 30.07 | 0.0168 | 29.97 | 0.0166 | 29.56 | 0.0156 |
| Verbs | 14.46 | -0.0155 | 15.07 | -0.0144 | 16.00 | -0.0127 | 15.94 | -0.0128 |
| Adjectives | 6.21 | Compliant | 6.31 | Compliant | 6.49 | Compliant | 5.94 | Compliant |
| Pronouns | 7.66 | Compliant | 8.73 | Compliant | 9.40 | Compliant | 8.70 | Compliant |
| Adverbs | 4.33 | -0.0123 | 4.21 | -0.0130 | 4.35 | -0.0122 | 3.72 | -0.0161 |
| Numerals | 0.77 | -0.0292 | 1.04 | -0.0257 | 0.65 | -0.0308 | 0.86 | -0.0280 |
| Conjunctions | 5.45 | 0.0357 | 5.33 | 0.0338 | 7.34 | 0.0645 | 6.82 | 0.0566 |
| Prepositions | 9.89 | 0.0548 | 9.13 | 0.0472 | 8.12 | 0.0370 | 7.40 | 0.0298 |
| Particles | 9.66 | 0.0061 | 8.19 | Compliant | 10.11 | 0.0085 | 10.95 | 0.0130 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Model Particles | 0.31 | -0.0117 | 0.10 | -0.0225 | 0.15 | -0.0198 | 0.15 | -0.0202 |
| Word Length | 4.20 | 0.0691 | 4.22 | 0.0696 | 4.49 | 0.0772 | 4.31 | 0.0722 |
| Top 30 High-Frequency Words | 43.56 | 0.0082 | 38.65 | 0.0024 | 41.07 | 0.0053 | 44.09 | 0.0089 |
| Average Sentence Length | 16.26 | Compliant | 16.46 | Compliant | 14.18 | Compliant | 14.92 | Compliant |
| Exclamatory Sentences | 0.18 | Compliant | 0.08 | -0.0062 | 0.36 | 0.0099 | 0.47 | 0.0323 |
| Periods | 4.97 | 0.0091 | 5.30 | 0.0127 | 5.26 | 0.0123 | 4.76 | 0.0067 |
| Question Marks | 1.05 | 0.0586 | 1.14 | 0.0681 | 1.45 | 0.1026 | 1.51 | 0.1086 |
| Commas | 6.92 | 0.0086 | 6.49 | 0.0052 | 9.53 | 0.0287 | 10.14 | 0.0335 |
| Semicolons | 0.42 | 0.1145 | 1.09 | 0.4071 | 0.47 | 0.1341 | 0.88 | 0.3151 |
| Total Deviation Value | | 0.4460 | | 0.7448 | | 0.5724 | | 0.7721 |

In terms of the total deviation values, although these four translations do not exhibit significant disparities in overall quality, a meticulous analysis reveals that D. C. Lau's translation ranks first in terms of superior quality, followed by the Wenxin Yiyan translation, and then by Ezra Pound's translation and the Zhipu Qingyan translation. It is noteworthy that not only do translations by different translators vary in quality, but even among AI-generated translations, there exist marked differences in quality, thereby resulting in varying overall quality levels.

## 4. A Neural-Network-Based "Post-Editing" Approach to Enhancing Translation Quality

This study conducts an in-depth exploration of the strengths and limitations of human and AI translation at both the lexical and syntactic levels. Against the backdrop of the prevailing era of machine translation, a series of human-machine collaborative translation models such as "post-editing" have emerged. Achieving the organic integration of human and AI translation is not only key to enhancing translation quality but also an imperative path for promoting the sustainable development of the translation industry. In light of this, this study proposes a series of improvement pathways aimed at optimizing human-machine collaborative translation.

### 4.1. Improvement Pathways for AI Translation

Although AI translation currently represents the most advanced machine translation methodology, the research findings indicate that it still exhibits deficiencies in lexical selection, word-class conversion, sentence structure processing, and comprehension of the original semantic content. Enhancing the quality of AI translation constitutes one of the focal points for optimizing human-machine collaborative translation. Based on this, the following improvement pathways are proposed: First, it is necessary to optimize the algorithms and architecture of large language models by employing more advanced deep learning techniques and neural network architectures to enhance their language comprehension and generation capabilities, thereby improving the accuracy and fluency of translation. Second, the corpora of large language models should be expanded and optimized by collecting more high-quality, diversified corpus data covering texts from different domains, styles, and levels of difficulty, so as to enhance the machine translation system's learning and adaptation capabilities across various linguistic phenomena, and by cleaning and annotating the corpora to improve their quality and usability. Furthermore, the contextual comprehension capabilities of large language models need to be enhanced, with strengthened processing of complex sentences, metaphors, and cultural background information to avoid translation errors caused by insufficient contextual understanding. Finally, with regard to terminology translation in specialized domains, dedicated terminology databases should be established and effectively integrated with machine translation systems to improve translation accuracy in fields such as law, science and technology, and finance.

### 4.2. Improvement Pathways for Human Translation

As the core element in the human-machine collaborative translation process, the enhancement of translators' competence is of particular importance. The improvement pathways for human translation are as follows: First, translators' professional competence should be elevated through strengthened training in linguistic fundamentals, cultural knowledge, and domain-specific expertise, thereby enhancing their sensitivity to and comprehension of language, as well as their grasp of different cultural backgrounds, so as to better proofread and optimize machine translation outputs. Second, translators' awareness and capacity for human-machine collaboration should be cultivated, familiarizing them with the use of machine translation tools, their strengths and limitations, and equipping them with collaborative skills such as effectively editing and revising machine translation outputs and leveraging machine translation to improve work efficiency. Finally, a rigorous quality control system should be established, with clearly defined quality standards and workflows for human translation, supervision and evaluation of translators' work, and assurance of translation quality, alongside a feedback mechanism enabling translators to promptly understand translation quality conditions and make improvements.

### 4.3. Optimization Pathways for Human-Machine Collaborative Translation Models

In the current context of the gradual expansion of AI translation and the unprecedented prevalence of human-machine collaborative translation, achieving the organic integration of AI and human translation is of paramount importance. The improvement pathways for human-machine interactive translation models are as follows: First, collaborative models should be optimized by selecting appropriate models according to translation tasks and requirements, such as machine-translation pre-processing, human post-processing, or human-machine interactive translation. For large-scale translation of simple texts, a machine-translation pre-processing model may be adopted, wherein machines first generate preliminary translations, which are then proofread and revised by humans; for complex, specialized texts, a human-machine interactive translation model should be employed, enabling real-time interaction and collaboration between translators and machines during the translation process. Second, dedicated human-machine collaborative translation tools and platforms should be developed, providing functions such as terminology management, translation memory, and real-time collaborative editing to enhance the efficiency and effectiveness of human-machine collaboration. Furthermore, communication and coordination between translators and machines should be strengthened, with clearly defined division of labor and responsibilities, and prompt resolution of issues arising during collaboration, for example, by using translators' feedback and evaluation of machine translation outputs to help machines continuously learn and optimize. Finally, standardized human-machine collaborative translation workflows should be formulated, with clearly defined tasks and requirements for each stage to ensure the smooth progress of translation work, and with continuous optimization and adjustment of workflows based on actual conditions to improve translation efficiency and quality.

## 5. Conclusion

This study constructs a self-built corpus of English translations of The Analects and employs tools such as ABBYY Aligner, Word Smith, CLAWS7, and Em Editor to extract the 19 linguistic quantitative features described above. Based on the translation quality assessment model, the deviation values for each linguistic quantitative feature are calculated. A comparative analysis of the similarities and differences in the 19 linguistic quantitative features between human and machine translations is conducted from the lexical and syntactic levels. Finally, by examining the total deviation values of each translation, an accurate and objective quantitative assessment of the translation quality of each version is performed.

The findings of this study are as follows: At the lexical level, translators exercise flexibility in lexical selection during the translation process, resulting in relatively higher lexical richness compared to AI translation. With regard to frequency differences across word classes, translators not only flexibly convert verbs into nouns but may also convert other word classes into nouns, whereas AI translation requires further optimization in word-class conversion functionality. In terms of conjunction usage, AI translations employ a substantial variety of conjunctions to tightly connect sentences and paragraphs, forming texts with clear logic and complete structure. This indicates that, compared to human translation, AI translation systems can more effectively capture and express the logical relations in the original text, thereby generating more accurate and coherent translations. At the syntactic level, AI translation makes relatively few modifications to the sentences themselves when processing texts, whereas human translations pursue formal equivalence between the translation and the original text at the sentence level, exercising flexible processing of sentence structure and meaning during translation. This demonstrates that AI translation still exhibits deficiencies in the identification and judgment of sentence semantics, with certain limitations in analyzing and processing sentence structure and meaning. In the use of various sentence types, AI translation tends to directly replicate the punctuation marks of the original text, exposing the deficiency of large language models in comprehensive sentence processing capabilities, with their translation approach often limited to sentence-by-sentence translation. In contrast, human translators, based on their in-depth understanding of the original meaning, render the text in more natural and appropriate language, ensuring the fluency and accuracy of the translation.

Based on the in-depth analysis of this study, we observe that both human translation and AI translation technologies exhibit their unique strengths and limitations in the field of classical translation. In light of this, in translation practice, translation practitioners should adopt a synergistic and complementary strategy, deeply integrating the efficiency of AI translation with the expertise of human translation, with the aim of achieving simultaneous enhancement of both translation efficiency and quality.

## Acknowledgements

This research is supported by the following projects:

Project of the Social Science Achievement Review Committee of Hunan Province: “Quality Assessment and Post-Editing of ChatGPT Translations of Chinese Classics Based on the BLEU Model: A Case Study of the Four Books” (No. XSP25YBZ166).

Key Project of the Education Department of Hunan Province: “Construction of a Paradigm for English Translation and Dissemination of Chinese Classics Based on Natural Language Processing: A Case Study of the Four Books” (No. 24A0293).

Project: “A Comparative Study of the Quality of ChatGPT Translation and Human Translation of Chinese Classics Based on a Translation Quality Assessment Model” (No. LXBZZ2024236).

## References

[1] Baker, M. Towards a methodology for investigating the style of a literary translator[J]. Target, 2000(2): 241-266.

[2] Cui Xiliang. Artificial intelligence: Opportunities and challenges for language teaching[J]. Chinese Teaching and Research, 2024(02): 20-29.

[3] Dai Guangrong and Zuo Shangjun. The application and research of corpora in translation quality assessment[J]. Foreign Language Teaching, 2021, 42(02): 92-96.

[4] Han Hongjian and Jiang Yue. A corpus-based comparative study of linguistic features in human and machine literary translations—A case study of three translations of Pride

and Prejudice[J]. Foreign Language Teaching, 2016, 37(05): 102-106.

[5] Hu Kaibao and Li Yi. A study on the characteristics of machine translation and its relationship with human translation[J]. Chinese Translators Journal, 2016, 37(5): 10-15.

[6] Liang Junying and Liu Yiguang. The advantages of human intelligence in translation ability—A corpus-based comparative study of human and machine translation[J]. Foreign Languages and Their Teaching, 2023(03): 74-84, 147-148.

[7] Olohan, M. Introducing Corpora in Translation Studies[M]. London: Routledge, 2004.

[8] Qi Yuling and Jiang Yue. The construction of a literary translation quality assessment model based on linguistic quantitative features[J]. Journal of Xidian University (Social Science Edition), 2016, 26(01): 84-92.

[9] Sun Lin. Reflections on translation quality assessment[J]. Shanghai Journal of Translators, 2023(05): 37-41.

[10] Wang Kefei and Hu Xianyao. A corpus-based study of lexical features in translated Chinese[J]. Chinese Translators Journal, 2008, 29(06): 16-21, 92.

[11] Wang Kefei and Qin Hongwu. A discussion on linguistic features in English-Chinese translation—A macro-analysis based on a parallel corpus[J]. Foreign Language Research, 2009(01): 102-105.

[12] Wu Guangjun. The progress, meta-evaluation, and development directions of contemporary Chinese and Western translation quality assessment models[J]. Foreign Language Research, 2007(04): 73-79.