

## *Retraction*

# **Retracted: Research on Intelligent Calibration of English Long Sentence Translation Based on Corpus**

### **Security and Communication Networks**

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] H. Qiu, "Research on Intelligent Calibration of English Long Sentence Translation Based on Corpus," *Security and Communication Networks*, vol. 2021, Article ID 5365915, 9 pages, 2021.

## Research Article

# Research on Intelligent Calibration of English Long Sentence Translation Based on Corpus

Haimei Qiu 

School of Foreign Languages, Yinchuan University of Energy, Yinchuan 750105, China

Correspondence should be addressed to Haimei Qiu; haimeiqiu66@zcmu.edu.cn

Received 13 August 2021; Revised 10 September 2021; Accepted 14 September 2021; Published 24 September 2021

Academic Editor: Jian Su

Copyright © 2021 Haimei Qiu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous promotion and development of the new curriculum reform, English teaching is becoming more practical and comprehensive. As an indispensable part of daily English topics, the use rate and scope of English attributive clauses are extensive. Moreover, due to English attributive provisions, the length of the whole English sentence will inevitably increase; therefore, we can accurately understand and translate sentences by mastering the translation and understanding details of attributive clauses. In addition, there are noticeable differences between English and Chinese in attributive clauses. Chinese will not add a variety of modifiers like English but will directly put them in front of them as attributives, so we should pay attention to this in translation. This process increases the difficulty of the English translation. Therefore, this paper proposes a Corpus-based intelligent calibration of English long sentence translation. Based on the construction of the English long sentence Corpus, an intelligent calibration algorithm for English long sentence translation is designed, and experiments verify the effectiveness of this method.

## 1. Introduction

The general sentence structure in English sentences is subject-predicate structure, which produces other basic sentence patterns, mainly including S + VI sentence pattern, S + linkv + P sentence pattern, S + VT + O sentence pattern, S + VT + IO + do sentence pattern, and S + VT + O + >C sentence pattern, which form the basis of English sentences in turn. In long English sentences, these basic English sentence patterns often appear as the subject of sentences. Then, with the help of various forms of connective means, phrases, clauses, and additional components in English are added to the English main sentence, and simple English sentences are continuously expanded and extended to form long English sentences with complex structure [1]. From the analysis of the “shape” of English long sentences, we can vividly compare English long sentences to a lush tree with rich branches and branches, which is very complex. Therefore, after understanding the characteristics of the English language through this form, we can simplify the complex long English sentences layer by layer, separate the subject sentence in the long sentence, clarify the main

meaning of the long English sentence, ensure the accuracy of the translation of the subject of the long English sentence, and then add the additional content of the subject sentence, and finally translate the long English sentence correctly [2].

In order to solve the disadvantages of semantic ambiguity, inaccurate quantifiers, and low translation accuracy in the traditional translation methods with grammatical type variables, improve the quality of English language translation, and obtain a more accurate and smooth translation, Wang [3] proposed an accurate translation method of English language based on semantic analysis and Corpus and dictionary, grammar analysis, semantic analysis, and visual representation to comprehensively analyze the semantics of the English language and then use the translation template with syntactic type variables based on semantic analysis to translate the English language. According to the semantic analysis and combined with the grammatical type limiting conditions of the variables, he ensured that the words of the replacement variables meet the syntactic meaning types of the variables. The experimental results show that the proposed method can avoid semantic understanding ambiguity and improve the accuracy of English language translation.

However, this method only improves the translation accuracy to a certain extent, but it is for in-depth analysis of the calibration method. Gao and Chen [4], in order to improve the accuracy and reliability of memory-assisted long character English automatic translation, proposed a design method of an English automatic translation system based on the B/S framework to optimize and improve the system. Firstly, the automatic translation mode is introduced in detail, and then, the system's overall design is realized based on the system operation process, translation optimization algorithm, network topology, and software. Finally, the experimental platform is constructed based on the WinCC6.0 environment to test the system's overall performance. The results show that the designed English automatic translation system can effectively and quickly realize the memory-assisted long character English intelligent translation, with high data recall rate, good accuracy and reliability, and good system compatibility. However, the above two methods only improve the translation accuracy to a certain extent, but for in-depth analysis of the calibration method.

The main research contributions of the thesis include the following:

- (1) This paper proposes a Corpus-based intelligent calibration of English long sentence translation
- (2) An intelligent calibration algorithm for English long sentence translation is designed based on the construction of the English long sentence Corpus
- (3) Because of the use of English attributive clauses, the length of the entire English sentence will inevitably increase, so only by mastering the translation and understanding of the details of the attributive clauses can we understand and translate the sentence accurately. Experiments verify the effectiveness of this method

## 2. Definition of an English Long Statement

Comparatively speaking, long sentences have a complex structure and a large amount of information, suitable for expressing complex ideas. According to the length, sentences are divided into three categories: short sentences (1~9 words), medium-length sentences (10~25 words), and long sentences (more than 25 words). In the constructed corpora, such as the FLOB Corpus, the average sentence length of the English text is 26.26 words, and in the Brown Corpus, the average sentence length of the English text is 32.48 words, which has reached the definition of the length of long sentences [5]. Taking long news sentences as an example, in the book news reporting and writing, it is considered that short sentences contain insufficient information and are easy to cause ambiguity, while long news sentences are difficult to understand, so it is best to keep the sentence of news introduction within 35 words. The average length of each sentence in news reports is about 28 words, and the length of most sentences is 20~40 words. Based on the analysis of the above scholars, in order to facilitate the research, this paper takes the upper limit of its number and gives a working

definition for the relative concept of English long sentences, that is, in English texts, when the length of a sentence is greater than or equal to 40 words, such a sentence is an English long sentence.

## 3. Translation of Long English Statements Based on the Corpus

Considering that some English long sentences are complicated in structure, difficult to distinguish word order changes, and mixed with appositive insertion and ellipsis, for these sentences, using the above translation methods alone is difficult to ensure the accuracy and smoothness of the translation. In this case, the translator should greatly adjust the information organization of the translation, comprehensively process the whole sentence according to the semantic relationship, logical relationship, and thematic relationship, and design an intelligent calibration method for English long sentence translation based on the above research.

Parallel Corpus is the integration of a language text and its translation in B language [6]. This Corpus contains a large number of real translation examples and the translator's knowledge and skills. At the same time, it can also be applied to comparative language research. The source of Corpus depends on the purpose of the study. If we want to study the translation of a specific aspect of language, we need to collect a large number of parallel corpora. The current English parallel Corpus is difficult to meet this research purpose due to copyright and Corpus. Therefore, according to the research purpose, this paper makes a self-made English long sentence parallel Corpus.

*3.1. Source of the Corpus.* This paper collects Corpus from the Internet. All the original Corpus are from the "OK reading network" (<http://www.okread.info/index.php>). According to the observation of this paper, most of the English texts of the website come from authoritative websites such as FT Chinese network and New York Times. Paragraph alignment has been realized between English and Chinese texts, and the Chinese translation is mainly full-text translation. This paper downloads 3850 texts from the "OK reading network" from January to August 2020. After preprocessing, sentence-level alignment, extracting long sentences, and proofreading, 5364 qualified English sentences and 8574 Chinese sentences are obtained. The total size of the Corpus is about 670000.

*3.2. Degree of Correspondence of English-Chinese Sentences.* In the Corpus, the basic information of the corresponding degree of English and Chinese sentences is shown in Table 1:

In order to better explain the relationship between sentence length and comparison between English and Chinese sentences, English long sentences are divided into three groups, namely, 40~44 words, 45~49 words, and 50 words or more. Sentence comparison is integrated into three groups, namely, 1:1, 1:2, and 1 to many (including 1:3, 1:4, and 1:5). A Chi-square test is carried out after the data are

TABLE 1: Source of corresponding Corpus for English-Chinese sentences.

		Frequency	Percentage	Effective percentage
Effective segment	1 : 1	3357	62.8	62.8
	1 : 2	1426	29.8	29.8
	1 : 3	267	14.7	14.7
	1 : 4	18	6.4	6.4
	1 : 5	3	3.1	3.1

input into SPSS. The statistical results are shown in Tables 2 and 3.

As shown from Tables 2 and 3, the corresponding proportion of 1 : 1 sentences is always the largest for different English sentences, indicating that the main translation conversion unit of long English sentences still sentences. However, this proportion is significantly lower than the corresponding sentence proportion of 1 : 1 in the English-Chinese translation part of Beiwai general Chinese-English corresponding Corpus, which shows that long English sentences should be regarded as a special language side in English texts because of their large number of words, rich information, and complex syntactic structure. Therefore, although the sentence correspondence ratio of 1 : 1 can reach about 60% in this Corpus, it is difficult to reach the 1 : 1 correspondence degree of 70% or 80% as in the whole practical discourse or literary discourse.

Because the chi-square test value is less than 0.05, it reaches a significant level, indicating that there is a significant difference in the frequency percentage of at least one sentence comparison between the three groups. On the "sentence comparison is 1 : 1," the frequency percentage of the three groups shows a decreasing trend. From the corrected standardized residual value in the cross table, it can be seen that there is a significant difference between the groups. The residual value of 40~44 groups is significantly higher than that of 50 words and above, which shows that when the sentence length is 40~44 words, compared with the group with a sentence length of 50 words and above, translators prefer to use a Chinese sentence to correspond to a long English sentence. There is no significant difference among the three groups in the "sentence comparison of 1 : 2." From the percentage of 28.9%, 29.1%, and 30.4% in their respective groups, the percentage of sentence pairs of 1 : 2 has little difference. Although the proportion is less than the ratio of sentence pairs of 1 : 1, it is significantly more than the ratio of sentence pairs of 1 : more, which means that a long sentence is divided into two sentences. It is a common strategy in English long sentence translation.

On the "sentence comparison is 1 to many," the frequency percentage of the three groups shows an increasing trend, and there is a significant difference. The residual value of 50 words and above is significantly higher than that of 40~44 words, which shows that, compared with the group with a sentence length of 40~44 words, when the sentence length is 50 words and above, the translator will give more consideration to the substantial reorganization of the original sentence, split an English long sentence into three or more Chinese sentences, and reconstruct the sentence organization form of the translation.

Simplification refers to the tendency of the translator to subconsciously simplify the language information of the source text in the target language text. In the form of expression, semantic simplification and grammatical simplification may exist. For long English sentences, grammatical simplification at the syntactic level shows that translators simplify English sentences with tree structure into Chinese linear structure and break sentences according to Chinese syntactic characteristics. When translating long English sentences, splitting a long sentence into two Chinese sentences is the main strategy to simplify the syntactic structure of the original text. With the increase of the length of English sentences, the syntactic simplification of translated texts becomes more and more apparent, especially the splitting of a long sentence into multiple sentences. This shows that due to the requirements of the hypotaxis characteristics of Chinese sentences, the translator needs to adjust the sentence group organization form of Chinese translation, which is embodied in the ways of adding small sentence structure, splitting relatively independent components of meaning, reorganizing meaning groups, transforming English long sentences into multiple Chinese sentences, and transforming obstruction into accessibility.

*3.3. English-Chinese Correspondence.* When English is translated into Chinese, it is generally translated into 1700 ~ 1800 Chinese characters per 1000 word pairs. If the number exceeds or falls below this number, it may be overtranslation or undertranslation. There are 242 571 English words in this Corpus, 413 114 Chinese words, the text ratio is 1000 : 1703, and the overall text correspondence ratio is between 1000 : 1700 and 1000 : 1800. From the text ratio, the translation belongs to the appropriate translation.

There are 5193 pairs of English and Chinese sentences in the Corpus, and the text ratio of English words to Chinese words is between 1 : 0.84 and 1 : 2.93. In order to facilitate statistics, this paper changes the direction of text ratio and uses SPSS for ratio statistics and percentile description.

According to the results of ratio statistics, since the median value (1.689) is near the mean value (1.704), it can be considered that the data are concentrated near the mean value as a whole. The statistical results of percentiles can further prove that 50% of English long sentences have a Chinese-English ratio of 1.52 : 1 to 1.86 : 1, and 80% have a Chinese-English ratio of 1.39 : 1 to 2.04 : 1. These data are close to the empirical range of the Chinese-English ratio of 1.7 : 1 to 1.8 : 1, which is basically near a standard deviation (0.26). It shows that nearly 80% of English long sentences are

TABLE 2: Cross tables compared with sentence length.

			Sentence length		
			40 ~ 44	45 ~ 49	50 and above
Sentence contrast	1 : 1	Count	1485	920	713
		In sentence length/%	64.9%	62.3%	59.2%
		Total/%	34.2%	18.1%	16.7%
	1 : 2	Adjustment residual	4.2	3.0	-3.5
		Count	753	436	416
		In sentence length/%	28.9%	29.1%	30.4%
	1 to many	Total/%	15.9%	8.0%	8.2%
		Adjustment residual	-1.6	-0.9	1.7
		Count	92	90	108
		In sentence length/%	3.2%	6.9%	7.8%
		Total/%	1.6%	1.2%	2.7%
		Adjustment residual	-3.8	3.9	4.5

TABLE 3: Card-square inspection.

	Value	df	Progressive sig. (bilateral)
Pearson chi-square	48.629 <sup>a</sup>	4	0.000
Likelihood ratio	47.456	4	0.000
N in valid cases	5193		

<sup>a</sup>Expected count of 0 cells (0.0%) is less than 5. The minimum expected count is 71.41.

translated into Chinese with an appropriate amount of words, and there is no obvious overtranslation or undertranslation. The large correspondence ratio between Chinese and English shows that the amount of text in the translation is redundant, and there is a phenomenon of translation manifestation. The purpose of manifesting is to clarify the translation information. From the perspective of content, manifesting is divided into semantic manifesting and syntactic manifesting.

In the process of translation, when the source text contains places with unclear meanings, some places are often explained in detail in the translation [7]. From a cognitive perspective, translators often express their understanding process in the translation [8, 9], and the explicit processing is simpler than looking for the corresponding expression. For example,

- (1) We describe a field experiment measuring the impact of bundling instantly gratifying but guilt-inducing “want” experiences (enjoying page-turner audiobooks) with valuable “should” behaviors providing delayed rewards (exercising),” wrote the economists Katherine Milkman, Julia Minson, and Kevin Volpp of Wharton in a 2013 research paper. “We describe a field experiment to measure the effect of combining “want to do” with “should do.” “What you want to do is something that is pleasant at the time but will cause guilt (listening to fascinating audiobooks), while what you should do is something valuable and will bring feedback in the future (exercise).”

The Chinese-English ratio of this sentence is about 2.93:1. There are both semantic and syntactic manifestations in the translation. At the syntactic level, the translator divides the direct speech of the

original text into two sentences and guides the split translation by repeating want and should to make the translation level clearer. At the semantic level, the translator explains bundling, providing delayed rewards and Wharton, which is helpful for readers’ understanding.

The low text correspondence ratio indicates specific implicit processing in the translation [10]. Implicit, as opposed to explicit, means that the meaning or information expressed by lexical means in the source language is implicit in the specific context in the target language [11]. The results show that sentence length is only a manifestation of sentence complexity. For long English sentences with clear sentence relationships and meaning, the translator can reduce their formalization degree according to the context and obtain the best meaning transmission effect with the least number of words. For example,

- (2) As a reminder, the price of a barrel of crude oil plunged by 50 percent in the space of a few months in 2014 after a subtle change of policy in Saudi Arabia, one of the world’s largest oil-producing nations. As a reminder, in 2014, after Saudi Arabia, one of the world’s largest oil producers, fine-tuned its policies, oil prices fell by 50% in just a few months. The ratio of Chinese to English in this sentence is about 0.95:1. The word order of the translation is adjusted according to the thinking order of Chinese, but there is no additional supplement in the information, but some contents are compressed. For example, the price of a barrel of crude oil is simply treated as “oil price,” which does not affect the understanding of the translation.

*3.4. Implement English Long Statement Translation Based on the Corpus.* The translation of English long sentences generally follows the procedures of understanding, expression, and revision. First, the grammatical structure of the original text is analyzed and then combined with semantic analysis. Then, appropriate adjustments were made according to the language characteristics, habits, and ways of expression of Chinese and the needs of meaning and

expression. Finally, the original text is revised [12, 13]. According to the procedure of English long sentence translation, it can be seen that the analysis of grammatical structure is the top priority of long sentence translation. Therefore, when discussing the translation of long sentences in domestic translation textbooks, most of them focus on the operation at the syntactic level. This paper makes statistics on the use frequency of four translation methods: sequential translation, variable order translation, clause translation, and comprehensive translation, as shown in Table 4:

As shown in Table 4, the sequential translation method is the most commonly used in English long sentences, nearly half of the total. The second is clause translation, variable order translation, and comprehensive translation.

Because the language of news style is plain and easy to understand, the purpose is to describe events' occurrence and development process. Although long sentences contain more information, in most cases, their clear thinking, reasonable arrangement, and popular language meet Chinese readers' reading habits and cognitive expectations. In translation conversion, the translator can often adopt the sequential translation method, follow the order of the original sentences, organize the translation from front to back, and effectively express the original author's intention.

In terms of clause translation, because English long sentences are usually used to express complex content, although they are unified in form, they are often not closely related in meaning and relationship between subject sentence and subordinate sentence, subordinate sentence, and phrase and word, and each has relative independence. Chinese, on the other hand, prefers short sentences with single and distinct levels [14]. This difference determines that in the translation of long English sentences, part of the content is extracted and processed, which can make the translation hierarchical and semantic coherence.

In terms of variable order translation, the rich connections in English long sentences make their word order relatively flexible. Chinese is characterized by parataxis, which is used to organize information according to the order of time or condition first and result later. This difference determines that the word order often needs to be changed in translation. When the sequential translation is not ideal, the translator needs to consider adjusting the word order. So far, Corpus-based English long sentence translation is realized, as shown in Figure 1.

#### 4. Realize the Intelligent Calibration of English Long Sentence Translation

Considering that some English long sentences are complicated in structure, difficult to distinguish word order changes, and mixed with appositive insertion and ellipsis, for these sentences, using the above translation methods alone is difficult to ensure the accuracy and smoothness of the translation. In this case, the translator should greatly adjust the information organization of the translation, comprehensively process the whole sentence according to the semantic

relationship, logical relationship, and thematic relationship, and design an intelligent calibration method for English long sentence translation based on the above research.

*4.1. Extract English Long Sentence Translation Learning Parameters.* In order to improve the accuracy of English translation and the stability and accuracy of output, firstly, the English long sentence translation learning parameters are extracted, and combined with the fuzzy parameter information fusion method, the English long sentence translation distribution elements are obtained as follows:

$$C_1(m, n) = \sum_{m=1} x_m^*(t) + x_{m+1}(t) \sum_{n=1} x_{n-1}^*(t) + x_n(t), \quad (1)$$

where “\*” represents complex conjugate,  $t$  represents optimization time, and  $m, n$  represent element coordinates. According to the distribution of English long sentence translation, the spatial grid clustering method is used for multiobjective feature classification and mining, and the multiobjective detection statistical feature is obtained as follows:

$$\hat{f}(m, n) = \frac{e^{(m-n)} - e^{(m+n)}}{c_{4s_i}}, \quad (2)$$

where  $c_{4s_i} = \text{cum}\{|s_i(t)|^4\}$  represents the characteristic quantity of multiobjective parameter distribution, and  $s_i$  represents the kurtosis of fuzzy information.

According to the results of the above formula and the multiparameter fusion method, the optimal control of the multiobjective algorithm is carried out, and the statistical characteristics of the parameters are as follows:

$$C_{4S} = \hat{f}(m, n) \sum_{L=1}^n c_{4s_L}, \quad (3)$$

where  $L$  represents the parameter optimization length. Combined with the association learning algorithm, the multiobjective related parameters  $\Phi$ ,  $\Omega$ , and  $\Lambda$  are, respectively,

$$\begin{aligned} \Phi &= C_{4S} \sum_{x=1} e_x, \\ \Omega &= C_{4S} \sum_{x=1} e_x^{2L}, \\ \Lambda &= C_{4S} \sum_{x=1} e_x^{2L}. \end{aligned} \quad (4)$$

According to the obtained multiobjective related parameter  $x$ , the optimal parameter configuration is carried out by using the multiobjective configuration method to obtain the fuzzy membership function:

$$X(i+1) = \sum_{I=1} p(i) \pm b^* |m - X(i)|^* \ln\left(\frac{1}{u}\right), \quad (5)$$

where  $a$ ,  $u$ , and  $m$  are random numbers between 0 and 1,  $p(i)$  is the optimal solution of multiobjective optimization,

TABLE 4: Statistics of the translation methods.

Translation methods	Frequency of use (times)	Percentage (%)
Sequential translation	2418	43.4
Variable order translation	975	18.6
Clause translation	1258	22.7
Comprehensive translation	732	15.3

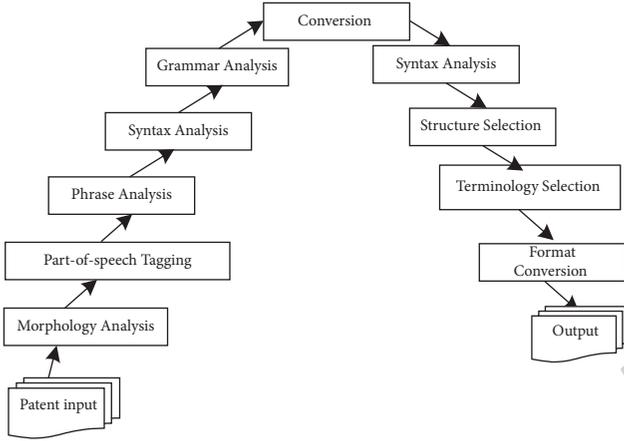


FIGURE 1: English long statement translation Corpus model.

that is, individual extreme value, and  $b$  is the multiobjective allocation coefficient.

According to the fuzzy membership function, the extracted learning parameters are

$$C' = \frac{C_{4S}A^{P-1}}{X(i+1)} + p(i) \cdot m, \quad (6)$$

where  $A$  is a multiobjective characteristic distribution matrix with dimension  $P \times L$ .

Based on the extracted multiobjective learning parameters, an English long sentence semantic ontology model is created.

**4.2. Creating Semantic Ontology Model of English Long Sentences.** English can be translated through the Corpus combined with the semantic features of English long sentences to realize the automatic calibration algorithm of the English translation. Therefore, the semantic ontology model of English automatic translation of English long sentences is constructed [15]. Suppose that the five-tuple  $E = \{C, I, H, R, D\}$  is applied to the semantics of English long sentences in English translation, and the fuzzy mapping of translation is set as

$$\begin{aligned} \theta: S &\longrightarrow S \times [-0.5, 0.5], \\ \theta(S_i) &= (S_i, 0), S_i \in S. \end{aligned} \quad (7)$$

The formula (7) represents the input factor of English long sentences and  $S_i$  represents the output function of English long sentences. The distribution structure model of English long sentences is defined as

$$\begin{aligned} A &\leq C, I, H, R, D, \\ A' &\leq C', I', H', R', D'. \end{aligned} \quad (8)$$

The concept of the English translation is realized by extracting the semantic features of English long sentences [16], and the parameters for automatic calibration of English translation can be obtained by using the method of fuzzy reasoning:

$$\Delta: [0, T] \longrightarrow S \times [-0.5, 0.5]. \quad (9)$$

This parameter can also be expressed as

$$\Delta(\alpha) = \begin{cases} s_k, & K = \text{round}(\alpha), \\ a_k = a - k, & a_k \in [-0.5, 0.5]. \end{cases} \quad (10)$$

In the formula,  $s_k$  represents the surface form of English long sentences,  $a_k$  represents the external context factor variables of English long sentences,  $\alpha$  represents the data set generated by the semantics of English long sentences, and the feature parameters of semantic fusion of binary English long sentences can be obtained by using the semantic mapping of associated English long sentences [17]:

$$(\bar{s}, \bar{a}) = \frac{\Delta \sum_{j=1}^n \alpha_j \alpha_j'}{\sum_{j=1}^n \alpha_j'}. \quad (11)$$

In the process of creating English translation, the evaluation index is set as  $I_j I$ , and an effective English long sentence semantic concept tree can be created through logical fuzzy reasoning, so as to obtain the model of English long sentence semantic ontology.

By evaluating the translation results of English machines through the semantic ontology model of English long sentences, the decision function can be obtained:

$$(\bar{s}, \bar{a})_{new} = \frac{\Delta \min_i \min_j \left[ (s_{ij}, a_{ij}), \alpha_j (s_j, a_j) + \max_i \max_j \Delta_{ij}^{-1} \right]}{(\bar{s}, \bar{a})}. \quad (12)$$

The long sentence translation design of the English long sentence semantic ontology model is realized through the above formula, so as to improve the semantic fuzzy matching ability of English long sentences for English translation error calibration.

**4.3. Long Sentence Translation Combination Calibration Algorithm.** Based on the analysis of the semantic features of English long sentences and the combination of long sentence translation [18], the automatic calibration algorithm of English long sentence translation errors is optimized, and the long sentence translation combination is created.

Dimensionless processing is carried out through the combination of long sentence translations [19, 20], and the English translation is automatically matched by the proximity estimation and the semantic similarity of English long sentences, so that the formula for calculating the comprehensive evaluation value of English translation output can be obtained:

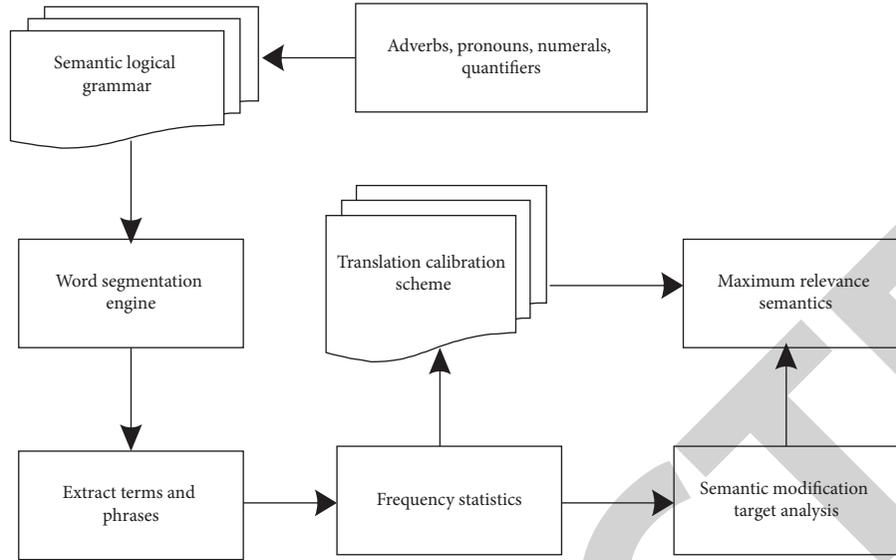


FIGURE 2: Implementation process of translation intelligent calibration.

$$Q_i = x(T_i + Y_i) + y. \quad (13)$$

In the formula,  $T_i$  refers to the relative closeness of distance,  $Y_i$  refers to the relative closeness of the combined relevance of long sentence translation, and  $x$  and  $y$  can fully reflect the relevance coefficient and comprehensive evaluation coefficient in English translation. The automatic calibration algorithm of semantic translation error combination of English long sentences is designed through the above algorithm. The implementation process is shown in Figure 2.

## 5. Test Analysis

In order to verify the application performance of this method, Matlab is used for the simulation test, and 30 tests are used for convergence analysis. It is assumed that the dimension of the semantic ontology vector is 24, the correlation coefficient is 0.21, the number of global iterations is 1000, the length of semantic information sampling is 1024, the similarity coefficient is 0.14, the correlation coefficient is 0.87, and the semantic correlation value is 1.24. It is set according to the above parameters, Design test indicators.

### 5.1. Test Indicators

**5.1.1. BLEU.** BLEU can calculate a proportion of the similarity of  $N$  groups of words between the comparative translation and the reference translation. Its value range is between 0.0 and 1.0. If the two sentences match perfectly, BLEU is 1.0. On the contrary, if the two sentences do not match, BLEU is 0.0. It can be seen that the closer to 1.0, the higher the translation quality and the more accurate the interaction.

**5.1.2. Root Mean Square Error.** The root mean square error can better reflect the ability of different methods to recognize long sentences in text. The smaller the root mean square error is, the better the recognition ability is.

**5.1.3. Calibration Accuracy.** The calibration accuracy will translate the preset English long sentence text, and each method will calibrate itself after translation. The translation quality of a translated paper will be evaluated through calibration accuracy.

**5.1.4. Calibration Speed.** The calibration speed is used to evaluate the speed of the system translating a paper. The faster the speed, the more timely the interaction.

**5.2. Analysis of Test Results.** According to the above parameter settings, in order to ensure the objectivity of the experiment, the accurate English language translation method based on semantic analysis in [3] and the automatic English translation system based on the B/S framework in [4] are selected as comparison methods to test the BLEU value, root mean square error, calibration accuracy, and calibration speed of different methods respectively.

**5.2.1. BLEU Index Test.** In Table 5, 1 and 2, 3 and 4, 5 and 6 correspond to two different types of texts from easy to difficult. As can be seen from Table 5, using different systems to translate the three test samples, it is concluded that the BLEU index value of the designed system is larger and the calibration speed is faster, which solves the problems of translation lag and error of the current translation system and improves the interaction between users.

**5.2.2. Root Mean Square Error Test Results.** In order to verify the effectiveness of Corpus-based intelligent calibration of English long sentence translation, an English text with 120 long sentences of the same difficulty is taken as the research object, and the methods proposed in this paper, [3, 4] are used to translate it, respectively. In order to ensure the objectivity of the experiment, the number of experiments is set to 5, and the root mean square error RMSEA of the

TABLE 5: BLEU index test results.

Experimental serial number	This paper is the method BLEU value	Reference [3] method BLEU value	Reference [4] method BLEU value
1	0.98	0.92	0.85
2	0.97	0.92	0.90
3	0.99	0.93	0.88
4	0.97	0.91	0.87
5	0.98	0.90	0.86
6	0.98	0.93	0.88

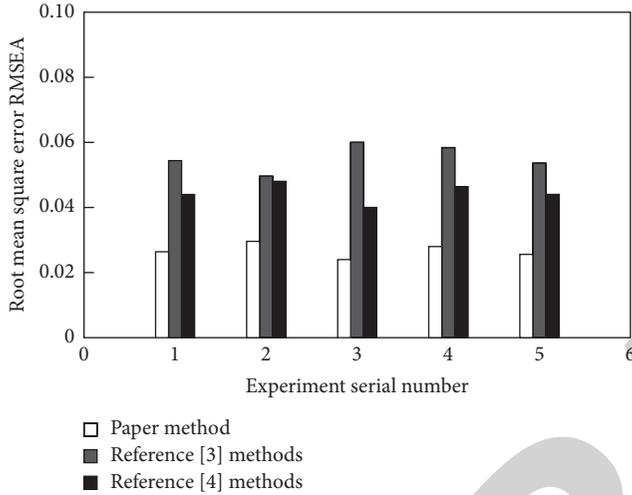


FIGURE 3: The RMS error test results.

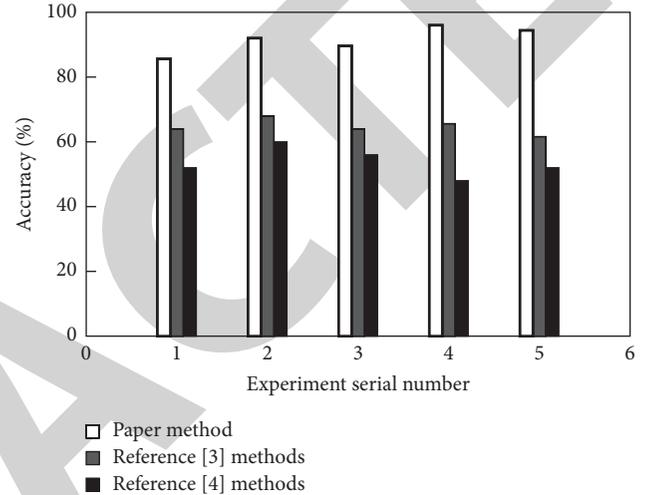


FIGURE 4: Calibration accuracy test results.

calculation results of different methods is compared. The test results are shown in Figure 3.

It can be seen from Figure 3 that the results obtained by the methods of [3] and [4] have large errors. The root mean square error obtained by the proposed method when translating long English sentences is less than 0.04, and the root mean square error is less than 0.04, which is within the acceptable range, resulting in less impact.

**5.2.3. Calibration Accuracy Test Results.** Take the calibration accuracy as the test index to test the proposed method, the method in [3], and the method in [4]. The calibration accuracy test results are shown in Figure 4.

It can be seen from Figure 4 that the calibration accuracy of the proposed method in many experiments is higher than that of the methods in [3] and [4], because the proposed method compares and analyzes English long sentences according to the designed Corpus and designs a calibration algorithm to improve the calibration accuracy of the proposed method.

**5.2.4. Calibration Speed Test Results.** The proposed method, the method in [3], and the method in [4] are used for testing, and the translation efficiency of different methods is verified by calibration speed. The test results are shown in Figure 5.

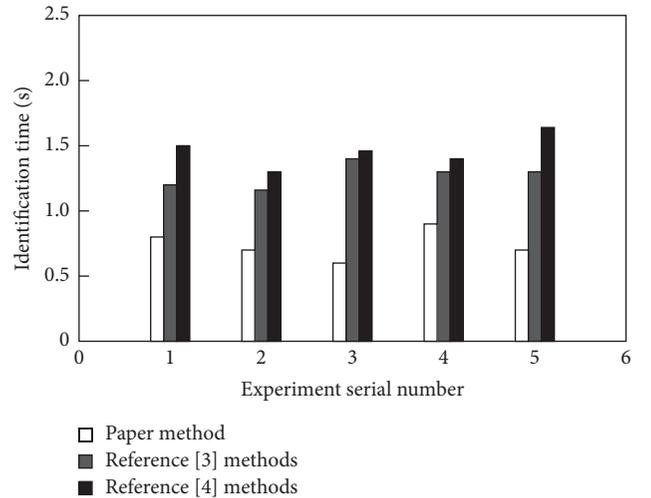


FIGURE 5: Calibration speed test results.

It can be seen from Figure 5 that the proposed method takes the least time to identify long English sentences, which proves that the proposed method can translate long English sentences in a short time, realize the translation of long English sentences, shorten the recognition time, and improve the recognition efficiency of the proposed method.

## 6. Conclusion

In solving the problem of English long sentence translation, learners must also fully grasp the different thinking characteristics and thinking habits of the East and the West and choose the appropriate language expression. Therefore, in translation practice, translators should not only compare the differences between the two languages and master their own characteristics but also consciously use the thinking characteristics of the two languages to find the appropriate translation methods and solve the problems encountered in English long sentence translation. In addition, in the study of English long sentence translation, learners should pay attention to the sentence structure in the original English text, learn to analyze the long sentence from the characteristics of the English language, and then combine the Chinese expression and adopt appropriate translation methods to make the translation reflect the logic of the original text on the one hand. Although our approach has achieved good performance, the model needs to be improved in terms of time complexity. In future work, we will try to design a model with lower time complexity and higher precision.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

There is no fund support for this work.

## References

- [1] B. Maa, B. Ap, and B. Fb, "A comparative study of machine translation for multilingual sentence-level sentiment analysis," *Information Sciences*, vol. 512, pp. 1078–1102, 2020.
- [2] S. Xabier, P. Olatz, L. Gorka, and O. Maite, "Neural machine translation of clinical texts between long distance languages," *Journal of the American Medical Informatics Association*, vol. 26, no. 12, p. 12, 2019.
- [3] J. Wang, "Research on English language accurate translation method based on semantic analysis," *Modern Electronics Technique*, vol. 41, no. 5, pp. 178–181, 2018.
- [4] N. Gao and W. Chen, "Design of memory assisted long character English automatic translation system based on B/S framework," *Techniques of Automation and Applications*, vol. 39, no. 8, pp. 36–38+55, 2020.
- [5] S. Maruf, F. Saleh, and G. Haffari, "A survey on document-level neural machine translation," *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–36, 2021.
- [6] F. Jiang, Y. Chiba, T. Nose, and A. Ito, "Automatic assessment of English proficiency for Japanese learners without reference sentences based on deep neural network acoustic models - sciencedirect," *Speech Communication*, vol. 116, pp. 86–97, 2020.
- [7] E. Rusconi and R. Cubelli, "The making of a syndrome: the English translation of gerstmann's first report," *Cortex*, vol. 117, pp. 277–283, 2019.
- [8] D. Raynaud, S. Gessner, and B. Mota, "Andalò di Negro's De compositione astrolabii: a critical edition with English translation and notes," *Archive for History of Exact Sciences*, vol. 73, no. 6, pp. 551–617, 2019.
- [9] F. Zhang, "Application of data storage and information search in English translation corpus," *Wireless Networks*, pp. 1–11, 2021.
- [10] Y. Qu and H. Liu, "The basic problem of contemporary water resources environment research: translation and analysis from the English practice report on water resources," *Journal of Coastal Research*, vol. 104, no. sp1, 2020.
- [11] X. Xu, "Machine learning-based prediction of urban soil environment and corpus translation teaching," *Arabian Journal of Geosciences*, vol. 14, no. 11, p. 1031, 2021.
- [12] B. Premjith, M. A. Kumar, and K. P. Soman, "Neural machine translation system for English to indian language translation using mtil parallel Corpus," *Journal of Intelligent Systems*, vol. 28, 2019.
- [13] Y. A. Ashengo, R. T. Aga, and S. L. Abebe, "Context based machine translation with recurrent neural network for English–Amharic translation," *Machine Translation*, vol. 4, pp. 1–18, 2021.
- [14] S. Li and C. Qi, "Computer aided translation model simulation based on bilingual E-chunk," *Computer Simulation*, vol. 36, no. 12, pp. 345–348+352, 2019.
- [15] M. Kaity and V. Balakrishnan, "An integrated semi-automated framework for domain-based polarity words extraction from an unannotated non-English corpus," *The Journal of Supercomputing*, vol. 76, no. 12, pp. 9772–9799, 2020.
- [16] M. Singh, R. Kumar, and I. Chana, "Corpus based machine translation system with deep neural network for Sanskrit to Hindi translation," *Procedia Computer Science*, vol. 167, pp. 2534–2544, 2020.
- [17] N. Chatterjee and S. Gupta, "Efficient phrase table pruning for Hindi to English machine translation through syntactic and marker-based filtering and hybrid similarity measurement," *Natural Language Engineering*, vol. 25, no. 1, pp. 171–210, 2019.
- [18] I. Haneef, R. A. Nawab, E. U. Munir, and I. S. Bajwa, "Design and development of a large cross-lingual plagiarism corpus for Urdu-English language pair," *Entific Programming*, vol. 2019, Article ID 2962040, 11 pages, 2019.
- [19] Z. He, "Self-calibration system for pragmatic failure in English-Chinese translation based on big data," *International Journal of Applied Systemic Studies*, vol. 9, no. 2, p. 141, 2020.
- [20] G. C. de Andrade, A. de Paiva Oliveira, and A. Moreira, "Ontological semantic annotation of an English corpus through condition random fields," *Information*, vol. 10, no. 5, p. 171, 2019.