Algorithm Design of Translation Accuracy Correction for
English Translation Softwares

Weiwei Qu*
School of Foreign Languages, Dalian Polytechnic University, Dalian, Liaoning, 116034, China
*Corresponding author e-mail: quweiwei@dlpu.edu.cn

Abstract. With the development of economic globalization, English is needed in more and more different cases in people’s daily life, and so is English translation. Traditional English software translation mainly depends on machine, which on the one hand facilitates people’s life, while on the other hand it also has some disadvantages. For example, translation errors are characterized by iterative transmission, weak logic and low accuracy. Traditional machine translation has been unable to meet people’s demand for both speed and quality, so it can no longer meet people’s needs. Therefore, this paper puts forward the design of translation accuracy correction algorithm for English translation softwares. First of all, an in-depth research on the traditional English translation methods, such as lexical semantic translation and phrase translation, are conducted with the method of documentation. Then, a dependency tree to string model and log linear model are established. Finally, the BLEU value and NIST value of the three translation softwares as well as the translation accuracy before and after algorithm correction is compared and analyzed. The conclusion shows that the highest accuracy of English translation before correction is only 75.6%, while the lowest accuracy is as high as 98.7% after the algorithm is adopted in this paper. The difference of accuracy between the two shows that the effectiveness of the correction system in this paper makes an outstanding contribution.

Keywords: English Translation, Software Translation, Accuracy, Dependency Tree to String Model

1. Introduction
The algorithm used in traditional machine translation is mainly the pipeline translation method, which is to analyze the sentence structure, sentence composition and part of speech of the original sentence, and complete the translation task after understanding the complete syntactic structure [1]. This translation method is easy to cause the accumulation of error iterations and low accuracy [2]. With the advancement of science and technology, higher requirements have been placed on translation technology. This era brings both opportunities and challenges to translation technology [3].

Considering the complexity and high cost of the automatic translation method used to construct the system under normal circumstances, an automatic translation method from dependency tree to string
model is proposed, and a correction algorithm suitable for English translation is proposed. Compared with the traditional method, this method only needs to analyze the synthesis structure of the source language, which greatly reduces the complexity of system construction and effectively reduces the cost [4]. In order to improve the accuracy of translation, by introducing a combined model of Chinese word classification, partial vocabulary tags and syntactic analysis, the core errors of the source language in English translation can be reduced, and the accuracy of the translation output function can be improved [5].

The innovations of this article are: (1) The combination of qualitative research and quantitative research, to fully analyze the research data; (2) The combination of theoretical research and empirical research, based on the original English translation software, three kinds of translation software conduct empirical investigation on the results of correction algorithms.

2. English Translation Software Translation Accuracy Correction Algorithm Design Method

2.1. The Semantic Similarity Method of Vocabulary Based on Hownet

The value range of similarity is [0,1], and the semantic similarity between different words W1 and W2 is:

$$\text{Sim}_{\text{semantic}}(W_1, W_2) = \max_{i=1,2,\ldots,n; j=1,2,\ldots,m} \text{Sim}(S_{i1}, S_{2j})$$  \hspace{1cm} (1)

In formula (1), $S_n(i=1,2,\ldots,n)$ represents n concepts corresponding to vocabulary W1, and $S_m(i=1,2,\ldots,m)$ represents m concepts corresponding to vocabulary W2. The so-called semantic similarity is the value with the highest similarity among multiple concepts of these two words.

The semantic similarity of vocabulary concepts can be used to describe the similarity of vocabulary concepts. The similarity between Yiyuan p1 and p2 can be calculated by the following formula:

$$\text{Sim}(p_1, p_2) = \frac{\alpha}{d + \alpha}$$  \hspace{1cm} (2)

In formula (2), the adjustable parameter is used to indicate that the values of the two meanings are both greater than 0, and the distance between them is expressed by d.

2.2. Computer Intelligent Correction Method Based on Improved Phrase Translation Model

The conversion from one text format to another is the process of English translation correction. Therefore, the intelligent computer correction process of English translation is actually a process of translation again. By comparing the revised results with the original translation results, an English translation with higher accuracy can be obtained cleverly. In this article, H is defined as the wrong translation result of English, and D is defined as the correct English translation result. The conversion from H to D is the process of English translation. The English automatic translation method based on the improved phrase translation model is as follows:

$$\hat{D} = \arg \max_c M(D|H) = \arg \max_c M(H|D) \cdot M(D)$$  \hspace{1cm} (3)

In the formula, the proofread vocabulary is described by M(D).

The most important thing in English machine translation is to improve the translation accuracy of vocabulary. Therefore, the intelligent computer English translation correction system can fundamentally solve the problem of English translation accuracy. M(D) in formula (4) represents accuracy. Therefore, based on the optimization formula (3), the computer intelligent correction is realized, and the specific method is as follows:
\[ \hat{D} = \arg \max_c M(\hat{H} | \hat{D}) \cdot M(D) \]  

(4)

In the formula, the weights of \( \varepsilon, \gamma \), the weights of \( M(H|D) \) and \( M(D) \) are represented by \( c \) respectively.

In order to facilitate the expression of the computer intelligent correction method based on the improved phrase translation model, \( H \) represents the vocabulary to be corrected, and the corrected vocabulary is represented by \( D \). It is defined that \( p \) characters exist in \( H \), represented by \( H_p \); at the same time, \( q \) characters exist in \( D \), represented by \( D_q \). The definition divides \( H_p \) into \( d \) random strings, represented by \( \hat{H}_d \), where the strings correspond to the phrases in the phrase translation model. In the same way, the proofreading vocabulary generated by the proofreading vocabulary contains \( d \) strings, which are described by \( D_d \). In summary, the extended form of equation (4) is obtained as follows:

\[ \hat{D} = \arg \max_{D, M_d} \sum_{M_p} M(\hat{H}_d | H_p) \cdot M(\hat{D}_d | D_q) \cdot M(D) \]  

(5)

In the process of intelligent correction of English translation, it is necessary to focus on the translation skills and methods of vocabulary in different scenarios, and to proofread the translation results one by one, and finally carry out the correction process to improve the accuracy of English translation. Combining the method described by the above formula to find the vocabulary \( D \) corresponding to the vocabulary \( H \) to be proofread, realize the intelligent correction of English translation computer.

3. English Translation Software Translation Accuracy Correction Algorithm Design Model

3.1. Dependency Tree to String Model

The representation form of dependency tree to string model is \( <D, S, A> \). Here, \( <D, S> \) is the translation pair, \( D \) is the source language dependency tree, \( S \) is the target word string in the source language, and \( A \) is the relationship between \( D \) and \( S \). The source language dependency tree \( D \) of the word alignment relationship. Each \( D \) word has a characteristic. The English below each word represents the sound part of that word. For example, the noun is NN, the verb is VV, the adjective is JJ, etc. For the relationship between characters and words, the third part of the model has English \( S \) strings corresponding to Chinese sentences. The connection of the upper and lower parts can be used to illustrate the connection configuration between the Chinese word node and the English word [6].

3.2. Log-Linear Model

The log-linear model is a model that can be judged, and multiple features are selected to judge divergent thinking. For a set sentence \( f_i = f_i, \ldots, f_j \), the translation model \( e_j = e_1, \ldots, e_j \) is formed with the maximum entropy:

\[ e_j = \sum_{m=1}^{M} \lambda_m h_m(e_j, f_j) \]  

(6)

The log-linear model has strong scalability, can set characteristics corresponding to various target requirements, and can apply machine conversion methods in multiple languages. The main form of the machine translation system is the possibility of forward and reverse translation and the switching of translation language modes. According to the actual needs of the translation system, customize the function system and determine the weight of the corresponding authority, and obtain the highest translation score according to the above formula [7].
3.3. Probability Calculation
The formula for calculating the probability of English translation accuracy is as follows:

\[ P(s|t) = \frac{\text{count}(s,t)}{\sum_{s} \text{count}(s,t)} \quad (7) \]

In the above formula, the left side represents the forward translation probability of the source language phrase \( s \) translated into the target language \( t \), where \( s, t \) represents the number of occurrences in the translation [8].

4. English Translation Software Translation Accuracy Correction Algorithm Design Analysis

| Table 1. Numerical comparison of different English translation software |
|-----------------------------------------------|
| Translation software | BLEU value | NIST value |
|-----------------------|------------|------------|
| Google Translate      | 25.43      | 5.7673     |
| Baidu Translate       | 25.18      | 5.7318     |
| Netease translation   | 24.53      | 5.7624     |

The BLEU in the above table is a comparative analysis of the unit segment of the evaluation translation and the reference translation. The more corresponding segments, the higher the translation quality. NIST is a measurement standard for evaluating translation quality. Used to evaluate the translation quality of each unit. The higher the NIST value, the higher the quality of the translation. It can be seen from Table 1 and Figure 1 that the BLEU value and NIST value of Google translation software are higher than the other two translation software. This shows that the quality of the translated articles of Google Translator is higher [9].
Table 2. Comparison of translation accuracy before and after correction algorithm

| Experiment serial number | Translation accuracy | Before correction/% | After using the correction algorithm/% |
|--------------------------|----------------------|--------------------|----------------------------------------|
| 1                        |                      | 68.8               | 99.8                                   |
| 2                        |                      | 72.7               | 98.8                                   |
| 3                        |                      | 67.9               | 98.7                                   |
| 4                        |                      | 72.4               | 99.8                                   |
| 5                        |                      | 75.6               | 98.9                                   |
| Mean accuracy            |                      | 71.5               | 99.1                                   |

Figure 2. Comparison of translation accuracy before and after correction algorithm

It can be seen from Table 2 and Figure 2 that the highest accuracy of the English translation results before correction is only 75.6%. After the correction of this text system, the lowest accuracy is as high as 98.7%. The difference of accuracy between the two shows that the effectiveness of the correction system in this paper makes an outstanding contribution. From the perspective of the average translation accuracy, the average English translation result before uncorrected is only 71.5%. After the system is used for correction, the average accuracy is increased by 27.6%, which again verifies the effectiveness of the computerized intelligent correction system for English translation [10, 11].

5. Conclusion
This article mainly studies the design of the correction algorithm for the translation accuracy of English translation software. The correction algorithm is researched on the basis of the original translation software. The purpose is to effectively improve the translation accuracy of the English translation software. Experiments have proved that the accuracy of translation before correction is significantly lower than that after correction, indicating that the correction algorithm in this paper has played a very good effect. The innovation of this article is the combination of qualitative analysis and quantitative analysis, which is well represented in the fourth part. The disadvantage of this article is that the number of experimental subjects selected is limited and the experimental results need to be treated more rigorously.

Reference
[1] Lee, Vivian. Considerations of Relevance to the Target Reader in Korean into English
Translation [J]. Asia Pacific Translation & Intercultural Studies, 2016, 3(1):62-75.

[2] Calefato F, Lanubile F, Conte T , et al. Assessing the Impact of Real-Time Machine Translation on Multilingual Meetings in Global Software Projects [J]. Empirical Software Engineering, 2016, 21(3):1002-1034.

[3] Muravev Y. Teaching Legal English Translation by the Case Method in Russian-English Language Pair [J]. Humanities & Social Sciences Reviews, 2020, 8(4): 961-971.

[4] Zhao Q , Zhao Y , Cong C , et al. Correction of the Projection Center of Rotation Based on the Isogram Using Translation Matching Method [J]. Journal of Biomedical Engineering, 2018, 35(4):598-605.

[5] Liu Q, Mohy-Ud-Din H, Boutagy N E, et al. Fully Automatic Multi-atlas Segmentation of CTA for Partial Volume Correction in cardiac SPECT/CT [J]. Physics in Medicine & Biology, 2017, 62(10):3944-3957.

[6] Shai, Bergman, Mark, et al. ActivePointers: A Case for Software Address Translation on GPUs [J]. Computer architecture news, 2016, 44(3):596-608.

[7] Isabelle Duvernoy, Sylvie Paradis. Translation of speech into text.(Computers & Software)[J]. Cybergeo, 2016, 64(1):61–83.

[8] Rokicki S, Rohou E, Derrien S . Hybrid-DBT: Hardware/Software Dynamic Binary Translation Targeting VLIW [J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2019, 38(10):1872-1885.

[9] Jeong J, Son Y , Oh S. Intermediate Language Translation and Evaluation for Binary Code Software Weakness Analysis [J]. International Journal of Multimedia and Ubiquitous Engineering, 2017, 12(10):15-26.

[10] Ferreira R, Denver W, Pereira M , et al. A Dynamic Modulo Scheduling with Binary Translation: Loop optimization with software compatibility[J]. Journal of Signal Processing Systems, 2016, 85(1):1-22.

[11] Zhenjie Sun. A Study on the Educational Use of Statistical Package for the Social Sciences. International Journal of Frontiers in Engineering Technology (2019), Vol. 1, Issue 1: 20-29.