Curriculum and Teaching Methodology (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/curtm.2022.051006
ISSN 2616-2261 Vol. 5 Num. 10

A Probe into the Design of 3 Computer Aided English Translation Teaching in the Era of Big Data

Ruijie Gao[1,*]

[1]*School of Marxism, Sichuan Water Conservancy Vocational and Technical College, Chengdu, Sichuan, China*

Keywords: Trados software, computer-aided translation, translation memory, similarity matching

Abstract: Confronted with the problems of low quality and efficiency of manual translation within high cost in current scientific and technical English, A computer-assisted English translation system based on the auxiliary translation software--Trados is designed. With the purpose of improving translation efficiency and quality, while reducing translation costs, the system function modules are constituted by three modules--translation memory, translation retrieval module, and translation editing environment. Firstly, a translation memory model is established, and the input English is accurately searched and verified through the translation memory library, and then the translation retrieval module is used to match the corpus with higher similarity, specifically by using the "shortest edit distance" method to verify the corpus Perform evaluation calculations and classify corpora with a higher degree of similarity. After preliminary matching in the translation editing environment, a more realistic and accurate English corpus translation can be realized. Finally, the translation results are tested. The test results reveal that the Trados-based computer-assisted English translation system proposed in this study has high translation accuracy and professionalism, which can greatly improve the quality and efficiency of translation. This study is characterized by practical meaning for the development of translation teaching and research field.

1. Introduction

With the rapid development of big data and cloud computing, the earth has gradually developed into a "global village", and people all over the world can communicate beyond the limits of time and space through the Internet. The establishment and implementation of global economic integration has promoted cultural exchanges and economic cooperation in various countries, and language differences have become a stumbling block for international communication. Among them, most of the scientific and technical English in the academic circle rely on manual translation. The translation cost is high with low quality and efficiency, which cannot meet the current fast-growing translation market needs. Therefore, a high efficient and functional auxiliary translation software system is urgently required. A former scholar Xiao Zhiqing [1]proposed the study of CAT+MT+PE teacher-student collaborative translation model based on the concept of human-computer interaction, and Huang Saisai[2] proposed the application of self-built corpus in business English translation

teaching. However, the researches proposed by the above scholars all have certain limitations and fail to solve many problems in current English translation in a real sense. At present, there are four major translation softwares commonly used in the translation field, namely SDL Trados, DAVU, CAT and Trados, etc. The most popular and widely used translation software is Trados[3-5]. Compared with other translation software, the usage rate of this software is 78 %. The reason for the high usage rate is that the software is characterized by translation memory function, which can store the translated corpus and text data in the translation memory. After inputting similar corpora in the system, the memory can combine the original translation results to obtain more for accurate translation results. In addition, the translation and retrieval module of the software has powerful storage and retrieval functions, which can store the corpus samples in the memory indefinitely, thereby improving the quality and efficiency of scientific and technical English translation. This research combines the experience of the above scholars to establish a Trados-based computer-aided English translation system to improve the quality in the process of translation, but also highly accelerate the efficiency in teaching.

2. The overall framework of the system

As shown in figure 1, the Trados-based computer-aided technical English translation system is mainly made up by four parts, namely, User layer, Service layer, Computing layer and Storage layer. Among them, the User layer is mainly used for search interface, file upload, content viewing, online translation and file download. The Service layer mainly includes translation engine and search, and the search is divided into cross-domain, multi-attribute and IPC classification search. The Computing layer includes Map, Reduce, merge processing, separation processing and load balancing. In short, by means of cloud computing, the input corpus information is calculated and translated, and a series of merging, balancing and separation processing are completed to obtain translation information[6-7]. What's more, the storage layer is Mainly used for text storage, directory storage and index storage.

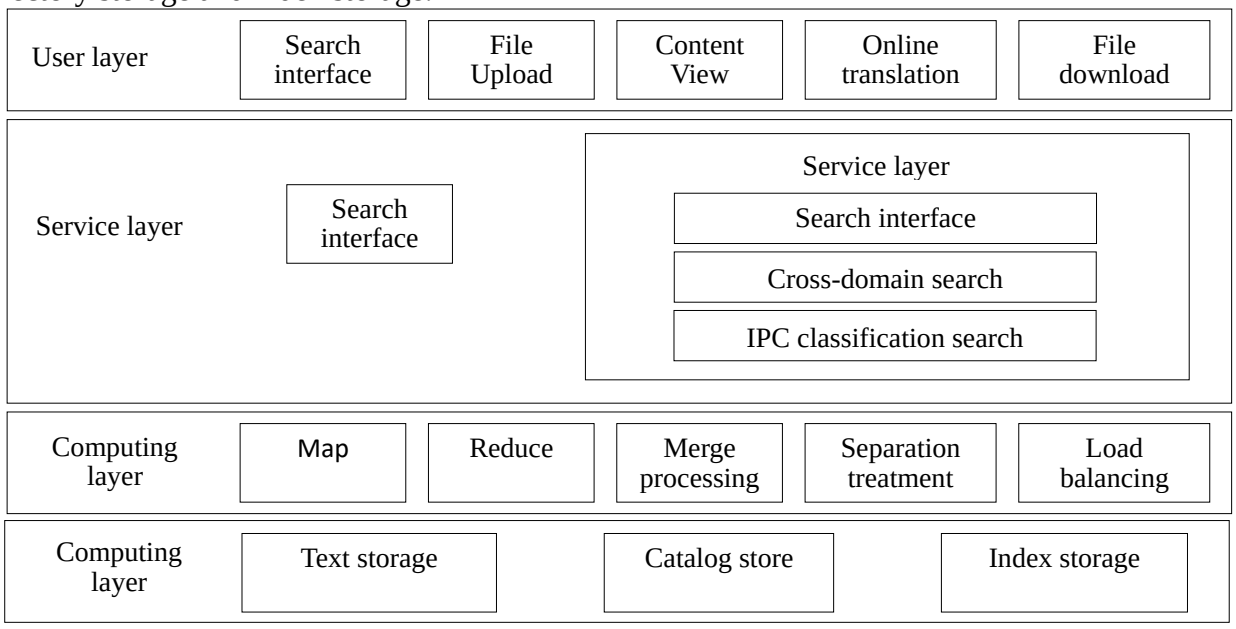

Figure 1: The overall framework of the system

3. System design and implementation

3.1 Main functional modules

The main functional modules of the system are divided into three parts, namely translation memory, translation retrieval module and translation editing environment. The functional modules are described in detail below.

3.1.1 Translation memory

The main function of the memory database is to store and memorize or extract the corpus text that has been translated and waiting to be translated. It belongs to the core functional module of the system. and the translation quality and efficiency of the system can be further improved through it. Due to the different types of memory databases, they can also be divided into sentence database and term database. When the system translates the input complex corpus, it can show the role of the translation memory, which can calculate the similarity of multiple corpora, retrieve the specific content of the complex corpus, translate and modify the input complex corpus. It can be seen that the data management and classification methods of the memory database directly affect the performance of the translation system. Therefore, the memory database must be planned and arranged in a reasonable and orderly manner. In Table 1, according to the unique sentence structure and vocabulary characteristics of Scientific English, the scientific and technical English sentences are arranged in an equal-level manner, classified and aligned to the same level. The specific structure is shown below.

Table 1: Translation memory structure

No.	Name	Description
1	Enter the sentence ID	Save all the entered scientific and technical English sentences in the same ID
2	Number of uses of the input sentence	Use the translation memory to record the number of times the input sentence is used in the database
3	The time the sentence was entered into the memory bank	Record the specific time and date when the sentence was entered through the translation memory
4	Sentence inputter's information	Record the specific information of the sentence inputter through the translation memory

Meanwhile, in addition to the above process, the design of the translation memory database must also retrieve the input English corpus information and add as well as delete it according to its semantics. Besides, the database should continuously update the sentences input by the system and extract the latest corpus information in time, so as to make sure that the database corpus can be kept in an active state.

3.1.2 Translation Search Module

The translation retrieval module is basically used as a reference template to simplify the redundant process in the course of translation and matching of scientific and technical English, thereby improving the quality and speed of complex translation procedure. To achieve this goal, the translation retrieval module is required as a standard. The detailed steps of the translation retrieval module are as follows. Firstly, search through the translation memory database is essential. After awakening a large-scale search, the historical input corpus text is obtained, and then the input corpus text is accurately searched and verified through the translation memory database. As the

module establishes a similarity matching algorithm, therefore, similar translation sentences can be matched through this module. By means of "shortest edit distance", the verified English data is estimated and similarity is divided by different degrees. After preliminary matching in the translation editing environment, a more precise and accurate English corpus translation can be achieved. It further shortens the redundant process of the complicated sentence structure and morphological translation of the previous scientific and technical English.

The similarity matching algorithm is the key to the retrieval mechanism. The first step of this research is to use the translation retrieval module to match the corpus with higher similarity, evaluate and calculate the verification corpus through the "shortest edit distance" method, and compare the degree of similarity. The high corpus is classified, and after preliminary matching in the translation editing environment, a more realistic and accurate English corpus translation can be realized. However, the sentences and grammar of scientific and technical English are very logical, and the semantic thinking is very meticulous. The sentence expression is professional and rigorous. These characteristics have raised the requirements and standards of English-assisted translation. It is essential to make the sentences translated in a precise way by the translation system in a higher context situation, otherwise it will not be able to accurately analyze the meaning of the English sentence. As a result, the difficulty of translation increases and the cost of manual translation increases. Based on this point, the Trados-based computer-aided technology English translation system designed in this research uses fuzzy matching algorithms to perform preliminary matching when the translation retrieval module searches and matches as well as judges sentences based on the similarity of the matching to improve English accuracy of translation. For example, set two sentences A and B, and calculate the similarity between the two, expressed as[8-10]:

$$Sim(A,B)=\frac{Max(A_{length},B_{length})-d[n,m]}{Max(A_{length},B_{length})}\times 100\% \tag{1}$$

In formula (1), A represents the target translation corpus sentence; B represents the historical reference translation in the translation memory. n and m indicate word frequency, $\boldsymbol{d[n,m]}$ indicates the distance of word frequency vector.

3.1.3 Translation and editing environment

The translation editing environment belongs to the frontier of the computer-aided technology translation system. Its function is to provide translators with data memory entries and filters inside the system, record and input format files, and the scientific and technical English segments and syntax that the system automatically divides and analyzes. Therefore, the translation editing environment can also be called a system pre-compilation or pre-processing module. The main function is to complete the preliminary work and proceed to the next step before the system performs retrieval and corpus matching, thus shortening the complicated steps of manual verification so as to improve translation efficiency. On the other hand, in the translation editing environment of the system, preliminary matching can be completed through a keyboard input. By means of the dichotomy, the existing vocabulary in the database can be searched, so that the similar English sentences or phrases can be automatically matched. In the next step, the translated sentences can be sorted by the system and be delivered into the hands of translators. After the translator revises and adjusts the received sentence, the translated English sentence can be exported in the format of the source document.

3.2 Overall system realization

The whole system needs to be realized through the C programming language, and the SQL Server database is used to store the corpus data to be translated in it. And then the bootstrap and Jquery two programs are used to construct the main operating interface of the system to achieve system visualization. As shown in Figure 2, the system process can be divided into three parts according to their different functions , dynamic matching segment of the memory bank, reference vocabulary and auxiliary reference translation. In the auxiliary reference translation part, the input English sentences can be edited and modified, and the translated paragraphs can be deleted. On the other hand, the translation memory can directly query complex segments and vocabulary, without having to find other programs for complex segment and vocabulary queries.

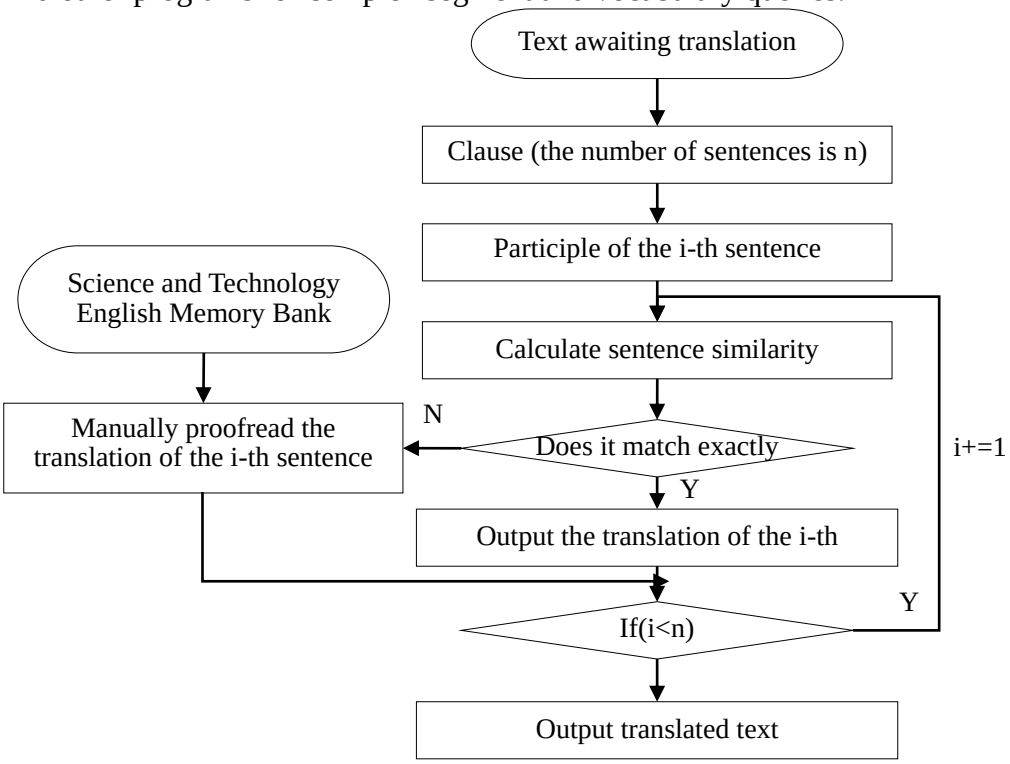

Figure 2: Flow chart of text translation

It can be seen from Figure 2 that the first step of the text translation process is to input the text to be translated, and then obtain the clauses of the n sentences and the i-th word segmentation. And then calculate the similarity of the input sentence, so the matching algorithm is adopted to compare the targeted sentence with similar data in the translation memory. The system will check the calculated result to see if it matches the sentence in the corpus completely. If the complete match is successful, the translation sentence of the i-th sentence can be output. Otherwise, you need to enter the manual verification part named the retrieval module in the Trados system, and the scientific and technical English memory library will be optimized by manual verification with the aim of getting accurate translation results. According to the term database possessed by the Trados system, each functional part needs managing, which can greatly improve the output productivity and matching accuracy of English translation and perfect teaching efficiency.

4. System application and test results

In order to test the system designed in this research, some corpus data will be randomly selected

from 1000 corpora as the test sample. The type of the test sample is the most commonly used form of scientific and technical English at present. The form of scientific and technical English is mainly classified as three types. They are test item A, test item B, and test item C respectively. The category of item A is tourism and finance, the category of item B is science and technology, and the category of item C is entertainment type. 6 corpus sample files with different numbers and different contents are randomly selected as follows. After filtering and classifying the selected samples, using the auxiliary translation system to perform segment morphological analysis on the input corpus samples. The results of the sample translation are obtained, and then the classification data of the sample translations are counted to realize the accuracy and precision of the translation corpus. The recall rate and other data, the sample translation classification results are summarized in Table 2.

Table 2: Sample translation classification data table

Test items	Classification results of each category (number of correct articles/wrong articles）							Accuracy	Recall rate	Accuracy
Test								/%	/%	/%
A	15/11	24/2	25/6	24/3	3/17	35/0	38/2	88.10	75.60	80.38
B	18T7	25/2	26/5	24/3	1/17	35/0	38/2	88.94	77.01	82.31
C	15/4	21/0	23/2	21/1	1/13	1/13	30/1	91.48	8138	86.77
Total	48/21 83.17	72/4	75/13	70/7	5/46	98/0	107/5	89.54	77.89	83.17

It can be seen from Table 2 that the classification accuracy rates of the three test sets are 88.1%, 88.94% and 91.48% respectively, the highest is about 91.48%, and the total classification accuracy rate is 89.54%. The recall rates of the three test sets are all above 75%, and the total recall rate is 77.89%. Compared with test item B and test item C, the relatively uncommon tourism and finance in test item A have the lowest average translation accuracy rate. But the lexical accuracy rate still remains above 80%, and the sum of the lexical accuracy rates of the three test sets It was 83.17%. In summary, the Trados-based computer-assisted English translation system proposed in this study is feasible and suitable for accurate translation of scientific and technical English. The final test results prove that the translation system features good applicability and sociability, improves the accuracy and professionalism of English translation, which is beneficial to the long-term business of current scientific and technical English translation.

References

[1] Xiao Zhiqing, Wei Guangfeng. (2021). Exploration of CAT+MT+PE teacher-student collaborative translation mode under the concept of human-computer interaction. Language Education, vol.9, no.2, pp: 69-74.

[2] Huang Saisai. (2020). The application of self-built corpus in business English translation teaching. Journal of Ningbo Radio & TV University, vol.18, no.3, pp: 117-121.

[3] Wang Jinfeng. (2019). Medical translation research from the perspective of computer-aided translation technology. Think Tank Times, vol.10, pp:188-189.

[4] Chen Zhenyun. (2019). Research on the strategy of integrating machine translation into college English teaching. Journal of Yancheng, Institute of Technology (Social Science Edition), vol.32, no.1, pp: 74-77.

[5] Wang Xue. (2019). Talking about the role of computer-assisted translation from the perspective of language service industry. Forestry Teaching, vol.5, pp:71-72.

[6] Li Xinyang. (2019). On the effective use of computer-aided translation in financial English translation. Chinese and foreign entrepreneurs, vol.21, pp:197.

[7] Zhang Jue. (2019). The adaptive development of English translation teaching under the "Internet+" education model. Journal of Jiangxi Vocational and Technical College of Electric Power, vol.32, no.5, pp: 46-47.

[8] He Liutao. (2018). Thoughts on the development of computer-aided translation in China under the background of artificial intelligence. English Square, vol.5, pp:17-19.

[9] Zhang Jian. (2019). Comparative study of TRADOS and Wordfast auxiliary translation tools. Journal of Ningbo

Institute of Education, vol.21, no.1, pp: 98-100.
[10] Liu Lin. (2020). Analysis of the advantages of Trados computer-aided translation software in translation practice——Taking the translation of public signs in the medical industry as an example. Journal of Hubei Open Vocational College, vol.33, no.1, pp: 174-175.