# English Translation proofreading System based on Information Technology: Construction of semantic Ontology Translation Model

## Rui Ma & Di Shao

Published online: 21 Apr 2023.

Submit your article to this journal ⤤

Article views: 371

View related articles ⤤

View Crossmark data ⤤

Taylor & Francis
Taylor & Francis Group

# English Translation proofreading System based on Information Technology: Construction of semantic Ontology Translation Model

Rui Ma and Di Shao

College of Foreign Languages, North China Institute of Aerospace Engineering, Langfang, Hebei, China

## ABSTRACT

Translation as a bridge between different cultures has received significant attention in recent years, especially with the rise of globalization and the increasing communication between nations. Translation has become a crucial tool for connecting people, promoting cross-cultural understanding, and facilitating international trade, academic exchange, and outbound tourism. With the advancement of information technology, research on improving English translation algorithms has become a hot topic. This is due to the complexities and variability of language, where words can have different meanings in different contexts. To address the issue of semantic coherence in current English translation proofreading systems, this paper proposes a semantic ontology translation model that utilizes fuzzy mapping to translate English and selects the most semantically appropriate translation using a decision function. The contributions of this research are twofold. Firstly, it provides a new approach to English translation proofreading by incorporating semantic ontology and fuzzy mapping. Secondly, it demonstrates the effectiveness of this approach through the improved efficiency and accuracy of English proofreading. The key findings of this study show that the proposed semantic ontology translation model can significantly improve the semantic coherence of English translation proofreading. This makes this article different from other studies by offering a unique solution to the problem of semantic coherence in English translation proofreading.

## Introduction

Since globalization, translation has become an important tool to mobilize people's relationships with others inspiring people to explore the unknown world and experience different cultures (Boonchieng et al. 2021). To improve the efficiency and accuracy of translation (Mamnunah and Mamnunha 2020) and meet people's needs for mutual communication

**CONTACT** Rui Ma ✉ mary1314aini@126.com ⊞ College of Foreign Languages, North China Institute of Aerospace Engineering, Langfang, Hebei, 065000, China

and understanding, machine translation has experienced a tortuous development. The first electronic computer came out in the 1940s and was suitable for machine translation. In the 1960s, the American Advisory Committee on Automatic Language Processing published the ALPAC report, which led to a two-decade slump in machine translation due to its low efficiency, poor accuracy, and much higher cost than traditional manual translation (Nazzal and Khmous 2021). The ALPAC report has been questioned and criticized by many as biased against machine translation, but it has done irreparable damage.

Machine translation has been widely used in every aspect of communication between countries and has become an important tool for communication between countries. In the mid-1970s, the study of machine translation entered people's businesses and became increasingly prosperous. However, in the early 1990s, the research of MACHINE translation fell into a low ebb again due to the complexity of natural language (Li 2020). In the 21st century, with the rapid development of information technology and the continuous improvement of algorithms, the quality of machine translation has been greatly improved (Mansor 2021). Machine translation is paid attention to again by all countries, and machine translation has ushered in a second upsurge (Wang 2021; Zhu 2021). The representative machine translation research projects (Deng 2021) in this period include the EURPOTRA Multi-language translation system of the European Community, the Taum-Meteo system of Canada, the Mu system of Japan, and the ODA program (Balanos-Medina 2012; Bo 2020).

This paper reviews and summarizes the development of machine translation and previous research achievements. It discusses the challenges of machine translation combined with professional background knowledge, the training of neural network machine translation systems, the accuracy of machine translation, and the verification process used to verify the English translation proofreading system. It concludes that machine translation has made a lot of breakthroughs, but there is still a lot of room for development. The system has high efficiency and accuracy and can meet the needs of most users for Proofreading English translations.

This study has two main contributions. It takes a fresh look at English translation proofreading by adding semantic ontology and fuzzy mapping. Also, it illustrates the efficacy of this strategy by increasing the efficiency and accuracy of English proofreading.

The rest of the article is outlined as follows:

1. Literature Review: This section provides an overview of previous studies and research related to the topic of English translation proofreading and the use of semantic ontology in translation.

2. English Translation Proofreading System: This Section Provides a Detailed Description of the Current English Translation Proofreading

System, Including Its Limitations and the Reasons for the Need for Improvement.

3. System Design: In this section, the authors present the design and implementation of the proposed semantic ontology translation model, including the use of fuzzy mapping and the decision function.

4. Results and Discussion: This section presents the results of the implementation of the proposed system and compares its performance with that of the current English translation proofreading system. The authors also provide a detailed discussion of the results and their implications.

5. Summary and Prospect: The final section of the article provides a summary of the main findings and contributions of the research, as well as a discussion of future directions for research in this area. The authors conclude by highlighting the potential impact of the proposed semantic ontology translation model on the field of English translation proofreading.

## Literature Review

The research content of this thesis belongs to the category of machine translation in computational linguistics. Here is a review of the previous research results and work of researchers in machine translation (Fan 2021; Tauchi 2020):

In 1949, (John-Jr 2012), inspired by language decryption and computers, formally proposed the idea of machine translation. Warren Weaver believed that the occurrence frequency, arrangement and combination of letters in different languages all had similar rules, which were similar to Morse code, so it could be used to translate languages automatically by machines. Warren Weaver proposed four possible implementation methods for machine translation: The first method is word substitution, which directly replaces words; The second method uses the logical expression of language to automatically deduce the translation between languages. The third method regards the automatic translation between languages as a process in which an input signal outputs another signal through a channel and restores the input signal according to the output signal. The fourth method makes use of the common law between languages. According to this law, an intermediate language is found as an intermediary. First, the language to be translated is converted into the intermediate language, and then the intermediate language is converted into the target language.

The first machine translation system was developed in 1954 by Sheridan of IBM and L. Dostert of the Computer Science Department at Georgetown University in Ohio. They carried out the machine translation experiment with an IBM-701 computer. The system translated 60 Russian sentences into English with good accuracy.

In 1973, R. F. Simon of The University of Texas established semantic Networks by using Fillmore's Case Grammar on the basis of Woods' ATN (Krüger 2016). Semantic network is used to express a complex concepts and their mutual relations of the directed graph, it is made of a semantic network composed of nodes and a directional arc described diagram to form, the nodes are used to represent all kinds of things, the concept and the situation, the properties, status, events and actions, etc., and with the direction of the curve indicates the semantics of the relationship between the nodes. Semantic network can not only describe the concept, situation, attribute and state of things themselves, but also describe the relationship between things. Semantic network represents knowledge based on basic semantic connections, which is also the basic unit of complex semantic connections. Therefore, it is possible to combine some basic semantic connections into arbitrary complex semantic connections.

In 1970, W. A. Woods of the United States put forward A natural language parser, Augment Transition Network (ATN), which makes use of context-free grammar for grammar analysis, and applied it to LUNAR system in 1972. When ANT starts working, the syntax of the input statements is parsed and identified, and when the parsed statements meet certain conditions, the statements are state transitioned and grouped in order. However, THE excessive dependence of ATN on syntax limits its ability to process some semantically compatible but not syntactic discourse.

In the early 1990s, statistical machine translation came into view. Peter F. Brown et al. of IBM (Miana et al. 2021) constructed a statistical machine translation model based on the idea of source channel, and carried out statistical analysis on a large number of parallel corpus, and then carried out translation. They describe five statistical models of a series of translation processes and give algorithms to estimate the parameters of these models given a set of mutually translated sentence pairs. The example they were given was limited to French and English translations, but they think the model can work well with other language pairs.

In 2003, YoshuaBengio (Brightman et al. 2021; Dirnagl et al. 2022; Doherty and Buckley 2021) and other scholars criticized the statistical machine translation model. They believed that the core of the statistical machine translation model is to carry out translation by constructing the joint probability function of word sequences in the language. As the number of words increases, the dimension of the joint probability function will also increase. It becomes very difficult to construct joint probability functions. To solve the problem of dimensionality, they propose to analyze and summarize the distribution of words instead of constructing joint probability functions, and propose a language model based on neural networks. That same year, Philipp Koehn (al), a professor in Johns Hopkins University's computer science department, proposed a new phrase-based translation model and decoding algorithm, and

he led the open sourceMoses machine translation tool that became a standard feature of the statistical machine translation era. When Google launched Google Translate in 2006, it used statistical machine translation.

Traditional statistical machine translation of words and phrases is ok, but the translation of long articles often appear incoherent sentences, sometimes even mistranslations. Dzmitry Bahdanau (He, Wu, and Li 2021), Yoshua Bengio and other scholars proposed neural network machine translation in 2014 to solve these problems. This translation system can maximize the overall performance of system translation by constructing a single neural network and adopting encoder-decoder framework for deep learning. They consider that the use of fixed-length vectors is the key to improve the performance of this basic architecture, and give relevant solutions. Then, they introduced the attentional mechanism of associating the original sequence elements with the output sequence elements into MACHINE translation, which further improved neural network machine translation greatly.

In NEURAL network machine translation decoding, beam search is usually used, which can find the local optimal target in each step of the search. However, due to the calculation of only one step forward, beam search usually cannot output the globally optimal target translation. In 2017, (Deng 2021; He, Wu, and Li 2021) and other scholars proposed the idea of using value network to improve neural machine translation for this problem (Galan-Manas 2011). This method adopts the circular structure of value network and uses bilingual data to train parameters bidirectionally. In the testing process, the neural network machine translation model will calculate the conditional probability function, and the value network will predict the long-term value. According to the conditional probability function and long-term value, decoding words can be obtained. Experimental results show that this method can significantly improve the accuracy of multi-language translation.

At present, unsupervised translation is the main direction of machine translation. Mikel Artetxe and other scholars put forward a new method to train neural network machine translation system in a completely unsupervised way (Feng et al. 2020; He, Wu, and Li 2021). This method does not require bi-directional training of parameters with bilingual data, but only needs mono-lingual corpus. The core of the unsupervised translation model is an improved Attentional encoder-decoder model, which uses a combination of denoising and translation to train parameters on a monolingual corpus. Although this method is much simpler than traditional neural network machine translation, in tests, the speed and accuracy of English-French and English-German translations are far superior to traditional neural network machine translation. When the translation content is limited to a certain field, the model can also use a small parallel corpus, and when the parameters are trained with 100,000 lines of sentences, the model also gets accurate translation results. This is a huge breakthrough in unsupervised neural network machine translation.

## English Translation Proofreading System

### Creating a Semantic Ontology Model

#### Semantic

Before constructing the semantic ontology model, the definition of semantics should be clearly defined. In essence, semantics are data. Data itself is just a symbol without any meaning. It can only be used by giving meaning to data and transforming it into information, and the meaning of data is semantics. Semantics can simply be thought of as the meaning of the concepts represented by the objects in the real world that data corresponds to and the relationships between those meanings, and the meaning of data is semantics (Ibragimovich 2021; Jimenez-Crespo 2012; Jimenez-Crespo, 2012, John-Jr 2012; Makhmudjonovna 2021).

Domain is the most important feature of semantics. The same thing represents different meanings in different fields. In terms of translation, the same word or phrase may have completely different meanings in different contexts.

Semantics are the data itself, while syntax is the definition of organizational rules and structural relationships between data. For the computer world, data and access to data are obtained by acting on schemas, where semantics refer to schema elements and syntax to the structure of schema elements (Feng et al. 2020).

#### Ontology

The definition of ontology was first proposed by The German scholar Studer: Ontology is a clear formal specification of a shared conceptual model. According to the definition of ontology, ontology should have the following four characteristics: first, ontology is a conceptual model, which is obtained by summarizing some universal phenomena in the objective world and reflects the common characteristics of substances; Secondly, ontology is unambiguous, and the concepts and conditions used in defining ontology are unambiguous. Thirdly, ontology is formalized. It should be written by programming language and can be modified or deleted by computer. Fourth, ontologies can be shared, not unique to individuals, because they reflect accepted theories in the relevant field (Krüger 2016; Maylath and Amant 2019; Vandepitte et al. 2016).

Ontology is built to define certain concepts and their relationships more clearly, so that people can better understand and share knowledge in related fields. In general, constructing ontology is of great significance for improving the ability of information transfer, interoperability and reliability of the system, and realizing knowledge sharing and application to a certain extent.

There is no unified standard about ontology construction method at present. However, in the process of constructing domain-specific ontology, it is universally acknowledged that experts in the domain are required to

participate. Five rules for constructing ontologies are generally accepted, as proposed by Gruber in 1995: First, an ontology must clearly state the meaning of the concepts it defines. All definitions should be stated in natural language. Definitions should be objective, complete and not subject to changes in the environment. Definition that can be described in the form of logical axioms; Second, an ontology should be consistent with the proof process that supports its definition. The axioms it defines and the related concepts it describes should be consistent; Thirdly, the ontology should be extensible. New properties may be discovered along with new research progress, so it should be able to define new concepts on the basis of existing concepts without modifying or recreating new ontologies. Fourthly, the construction of ontology should meet the minimum coding preference rule. Since the actual system may use different computer languages, the construction of ontology should not only use a specific coding method; Fifthly, when constructing ontology, the minimum rule of convention should be met, and the constraints on ontology should be minimized in the case of logic. In order to embody authority, experts in a specific domain are required to participate in the process of constructing ontology to meet the minimum rule of convention (Ren et al. 2022; Vandepitte et al. 2016).

### Semantic Ontology Translation Model

The translation process of traditional translation system is shown in the Figure 1. In traditional translation systems, translation memory banks need to be created first, which encode Chinese and English in XML language and store them in two libraries respectively. There is a mapping relationship between the symbols in the two libraries. According to the mapping relationship and the encoding of the input statement, the corresponding symbols can be found, and finally the symbols can be translated into translation. The key of a translation system is the matching between input statements and translation memory. However, due to the phenomenon of polysemy, the mapping relationship between symbols is one-to-many rather than one-to-one correspondence, so it is difficult for the system to judge the symbols in the current context (Wook-Dong 2021).
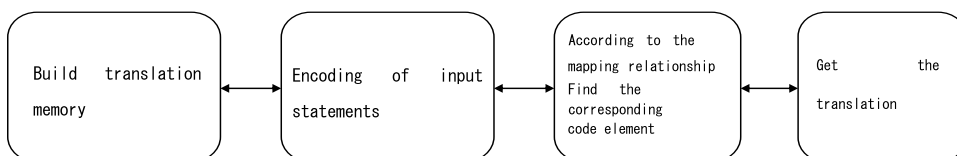


**Figure 1.** Traditional translation model.

**Figure 2.** Semantic ontology translation model.

The semantic ontology translation model can solve the problems of traditional machine translation. The structure of semantic ontology translation model is shown in Figure 2. First semantic ontology translation model can analyze the semantic features of input statement, and then fuzzy mapping, the fuzzy mapping to amplify the scope, the formation of a variety of translation, using the decision function to evaluate these English, choose the most conforms to the semantic translation, can be achieved English translation algorithm automatically, automatic translate English into Chinese.

The specific implementation process is as follows:

The input statement is first defined as an array, which, for illustration purposes, is assumed to be A five-element one-dimensional array G={A, C, H, I, R}.

The coefficient of fuzzy mapping determines the mapping range, the greater the coefficient, the greater the mapping range, and vice versa. If the coefficient is too small, matching translations may not be searched. If the coefficient is too large, too many translations will be produced, thus affecting the system speed. Here, the coefficient of fuzzy mapping is set as 0.5, and the fuzzy mapping of translation system is set as:

$$\theta : S \rightarrow S*[-0.5, 0.5] \qquad (1)$$

The range of fuzzy mapping being represented by [−0.5, 0.5] is a common representation in the field of fuzzy logic, where values within this range are used to represent the degree of membership of a given element in a fuzzy set. In the context of this research, the range of [−0.5, 0.5] is used to represent the degree of similarity between the source and target language words in the semantic ontology.

A value of −0.5 indicates that the source and target words have no similarity, while a value of 0.5 indicates that the two words are perfectly similar. Values in between −0.5 and 0.5 indicate varying degrees of similarity, with positive values indicating a higher degree of similarity and negative values indicating a lower degree of similarity.

This range is chosen because it provides a clear representation of the similarity between words, allowing the decision function to effectively select the most appropriate translation based on the calculated similarity values. Additionally, the range of [−0.5, 0.5] is widely used in fuzzy logic and is well

understood by researchers in the field, making it a convenient and intuitive choice for representing the degree of similarity in this research.

The frequency of English phrases varies. According to the distribution of English phrases, the structure model is defined as follows:

$$G \leq A, C, H, I, R \tag{2}$$

In the process of English translation, the input statements are divided into T segments, semantic features are extracted for analysis, and parameters of automatic English translation are obtained by fuzzy mapping method:

$$\Delta : [0, T] \rightarrow S*[-0.5, 0.5] \tag{3}$$

When the results of English machine translation are calibrated by semantic ontology model, decision functions should be used for judgment and evaluation:

$$(\xi, \eta) = \Delta \left( \frac{\min d(r, a) + \rho \max \Delta}{\Delta d(r, a) + \rho \max \Delta} \right) \tag{4}$$

The above equation is a decision function used in the evaluation of the English machine translation results, which have been calibrated by the semantic ontology model. The equation is used to determine the most semantically appropriate translation among the available options.

Here's what the different variables and symbols in the equation represent:

- $(\xi, \eta)$ represents the evaluation results of the decision function, which is a pair of values that indicate the degree of similarity between the source and target words in the semantic ontology.
- $\Delta$ represents the fuzzy mapping of the English machine translation results, which maps the input text to a set of target language words with a degree of similarity represented by the range $[-0.5, 0.5]$.
- $\min d(r, a)$ represents the minimum distance between the source and target words, which is calculated based on the semantic characteristics of the words in the semantic ontology.
- $\rho \max \Delta$ represents the maximum value of the fuzzy mapping $\Delta$, which represents the highest degree of similarity between the source and target words.
- $\Delta d(r, a)$ represents the change in distance between the source and target words, which is calculated as the difference between the minimum distance and the current distance between the words.

The decision function calculates the evaluation results as the ratio of the sum of the minimum distance and the maximum value of the fuzzy mapping, to the sum of the change in distance and the maximum value of the fuzzy mapping.

The evaluation results are used to rank the available translations and select the most semantically appropriate one as the final result.

Based on the above formulas, semantic ontology translation model is constructed to improve the semantic matching ability of translated English (Wright and Wright 1993).

### Phrasal Translation Combination Translation Algorithm

Rich structural forms usually pose challenges to machine translation. The difficulty of translating sentences and paragraphs is much higher than that of phrases and words. In the process of sentence translation, the sentence should be decomposed into phrases for translation, and combined with the analysis of semantic characteristics, the translation of phrases should be arranged and combined in a certain order to get the final translation result.

Dimensionless processing is carried out through phrase translation combination, and English translation is automatically matched by proximity estimation and semantic similarity, so as to obtain the calculation formula of comprehensive evaluation value Q of English translation output:

$$Q = xD + yR, x + y = 1 \tag{5}$$

There are two factors that influence the comprehensive evaluation value Q of English translation output: the relative closeness degree D of the input sentence and the translation, and the relative correlation degree R of the translation combination. D and R have different degrees of influence on Q, that is, they have different weights, which depends on the correlation coefficient X and comprehensive evaluation coefficient Y in the translation.

Automatic English translation can be realized through the algorithm mentioned above. If a client is added to the front end of the translation module, and a translation proofreading module is added to the back end, and combined with the software framework, the English translation system can be further optimized (Deng 2021).

The proofreading process of English translation is similar to that of English translation, which is to convert one language into another according to certain rules. English translation is to convert English into Chinese, and the proofreading process of English translation, in essence, is to convert the wrong part of the English translation into the correct part.

The wrong part of the English translation result is represented by W, while the correct English translation result, that is, the result after calibration, is represented by R. The process of converting W into R is the whole process of English translation proofreading. The calibration process of English translation can be expressed by Equation (3-6).

$$R = \arg maxM(R \vee M) = \arg maxM(M \vee R) \times M(R) \qquad (6)$$

In the formula, M(R) refers to translation accuracy. Since the model of this system is relatively simple, the accuracy of English word translation is still inadequate. Equation (3-6) can be optimized to improve the accuracy. In the process of English translation proofreading, the most important thing is to divide the wrong part of English translation results into appropriate lengths (Nazzal and Khmous 2021). Too long fragments will increase the difficulty of calibration, while too short fragments will lead to too many fragments, thus affecting the speed of calibration. The calibration module calibrates the segmented fragments and arranges them in the correct order, so as to obtain the final calibration result.

## System Design

The overall architecture of English translation proofreading system is shown in Figure 3. This English translation proofreading system consists of six parts, of which there are five basic modules, namely user module, search module, English translation module, English translation proofreading module and work module. The work data generated by user module, search module and work module are recorded by behavior log.

The five modules are responsible for different tasks. When the user login system, start the function, query results will use the user module; The search module is mainly responsible for analyzing the lexical characteristics and grammatical structure of input sentences (Wook-Dong 2021; Wright and Wright 1993). The English translation module translates the input statements into the target language. The English translation proofreading module is responsible for proofreading the translation results of the English translation module and replacing the errors in the original translation process;, work after school to complete the module will be received from translation module as a result, the work module analysis the characteristic of each word in the treatment of proofreading statements, according to the translation and the translation results sorted input statement similarity, finally choose the results accord with the actual translation sent to the user module, the user can search to the corresponding translation results (Balanos-Medina 2012).
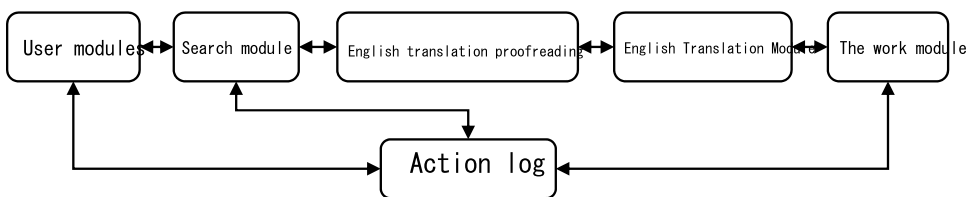


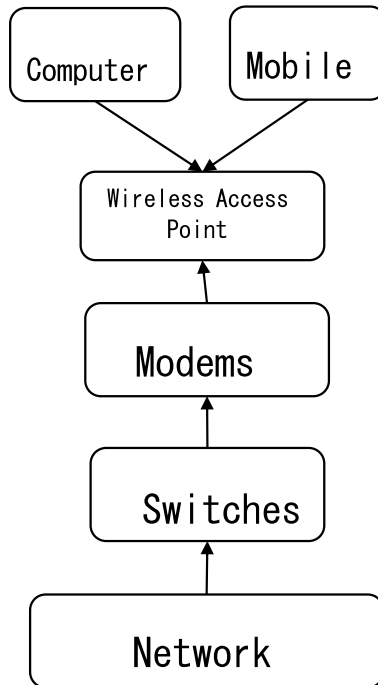**Figure 3.** English translation proofreading system.

**Figure 4.** Network topology.

In the process of the system using the above modules for English translation proofreading, the behavior data of each step of the user module, the search module and the work module will be recorded by the behavior log. Through the observation data, the background engineers can real-time understand the real situation of system work, timely discovery system in the working process of the existing problems, to solve these problems to develop targeted and effective measures, thus further optimizing the system performance of proofreading, improve the accuracy of the English translation proofreading system (Bo 2020).

The English translation proofreading system can be used online or locally. In the process of using the network, users can carry out English translation operations in the system, and can also check the update of the system. If the system has been updated, users will be reminded to download the latest version. Its network topology is shown in Figure 4. Hands and computers as clients can carry out wireless access, and the network translation server enters the wireless access point through switches and modems (Krüger 2016). In the process of local use, although online thesaurus and online search are not available, users can use the downloaded local thesaurus to realize local search, and users can also add words and modify the explanation of words in the local thesaurus according to their own needs, which is convenient for users to realize offline query of word information (de Boer et al. 2022).

## Results and Discussion

It is often difficult to compare the performance of various machine translation systems.

First of all, it is difficult to define the standard of translation results, whether it is machine translation or human translation "good," "medium," and "poor" quality levels. And there is no uniformity in the criteria for classifying each grade. Therefore, the evaluation of translation results often lacks objectivity and is difficult to be standardized.

Secondly, different users and occasions have different requirements for the results of machine translation. Some product manuals may require that the translation can be read and understood, while literary works may not only require accurate translation, but also require beautiful words and fluent phrases.

In addition, for different language structures (phrases, sentences, paragraphs, etc.), the efficiency of machine translation is also different. In comparison, the translation of phrases is the easiest, and it is not necessary to consider the preceding and following contexts, and it is sufficient to translate all the meanings of phrases, and the accuracy should be the highest. The translation of sentences and paragraphs, on the other hand, needs to be combined with the specific context, and it is relatively difficult to maintain the semantic coherence and smoothness of the articulation.The translation time required for phrases, sentences and paragraphs is also different.

In view of the above three reasons, the translation quality of the English translation system includes the comprehensibility and accuracy of the translation. The quality of the target translation is divided into three categories, A, B and C. The classification assessment guidelines are as follows.

Category A: The translation accurately expresses the meaning of the original text and is fluent and natural, or the translation accurately expresses the meaning of the original text and is fluent in its entirety, but is slightly inferior in grammar, choice of words for translation and Chinese expression habits, and can be revised to a fluent and natural translation without referring to the original text.

Category B: the translation can roughly express the meaning of the original text, and part of the translation differs from the original text, but it can be corrected without referring to the original text; or the translation can roughly express the meaning of the original text, and part of the translation differs from the original text, but it can be corrected by referring to the original text; or part of the translation conforms to the meaning of the original text, and the whole text is not translated correctly, but the words contained in the original text are translated in isolation, which is useful for the subsequent Human editing is of some use.

Category C: The meaning is completely incorrect, and we do not know what to do after reading the translation, but some parts or words are translated correctly, or no translation is done at all.

According to the above evaluation criteria, we randomly selected 100 phrases, sentences and paragraphs respectively as the corpus for the English translation system experiment, and evaluated the performance of the English translation model by testing the quality of the translations. The following test results were obtained from the English translation tests for phrases, sentences and paragraphs, respectively.

As shown in Figure 5, 96% of the test results for translating English phrases achieved category A, 3% achieved category B, and only 1% achieved category C. Except for the more remote phrases among them, all of them can be translated accurately. This English translation system basically works well for translating phrases.

As shown in Figure 6, among the test results of translating English sentences, 90% achieved category A, 7% achieved category B, and 3% achieved category C. All of them can be translated accurately except for individual phrases that are not fluent enough. The system's translation of sentences is slightly lower in quality than that of phrases, but the overall effect is good.

As shown in Figure 7, in the test results of translating English paragraphs, 82% of them reached Category A, 13% reached Category B, and 5% reached Category C. Although the quality of the system's translation of paragraphs is far inferior to that of phrases and paragraphs, it can meet the needs of most users overall.

In order to analyze the performance of the English translation proofreading module, the accuracy of the results of the English translation proofreading system before and after calibration is analyzed. As shown in Figure 8, before and after calibration, the accuracy rate of phrases increased from 95% to 99%, the accuracy rate of sentences increased from 92% to 97%, and the accuracy rate of paragraphs increased from 90% to 95%. It indicates that the English translation proofreading module has effectively improved the accuracy of translation.
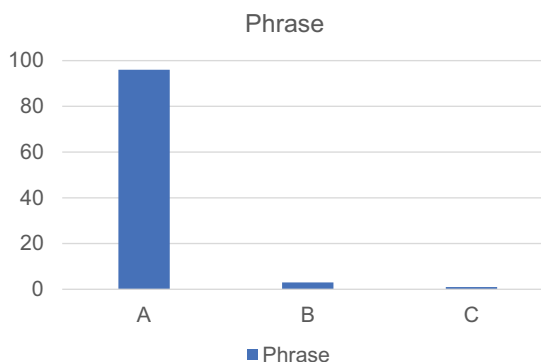


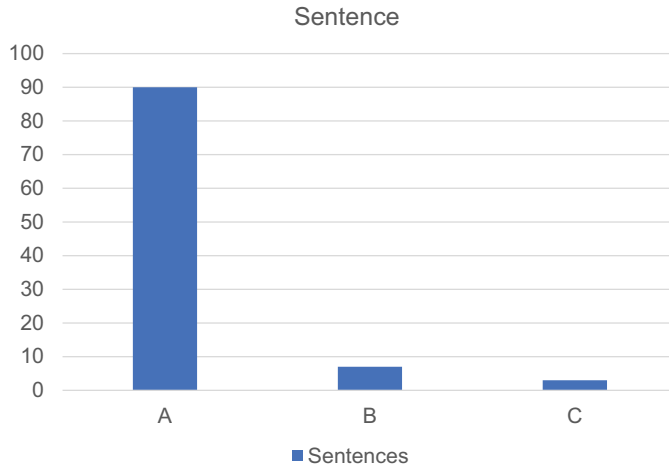**Figure 5.** Phrase translation result evaluation.

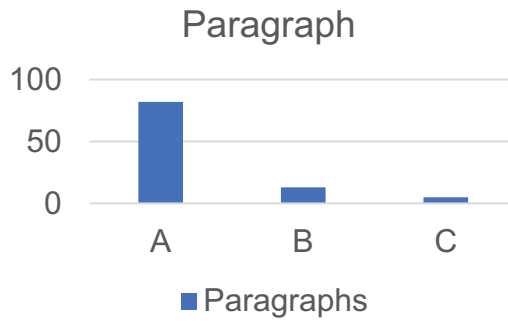**Figure 6.** Sentence translation result evaluation.



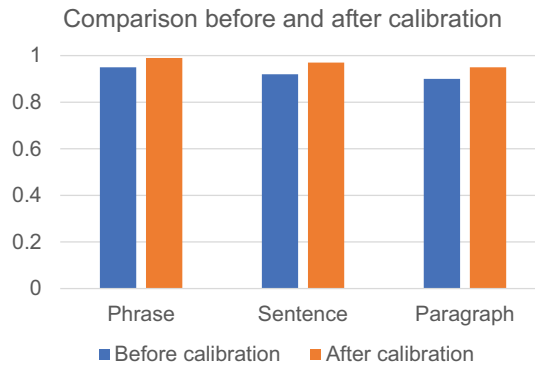**Figure 7.** Paragraph Translation Results Evaluation.



**Figure 8.** Comparison of results before and after calibration.

The preliminary experimental results show to a certain extent that the designed English machine translation proofreading system can basically translate English accurately and meet the needs of most users. However, there are still some errors in the translation process, and after analysis, the reasons for the errors in the translation are mainly the following: first, the scale of the current translation memory is too small, and some of the input words cannot be found in the translation memory and can be matched with the translation, thus a satisfactory translation cannot be obtained. Secondly, since the fuzzy mapping is not a one-to-one mapping, it inevitably causes local errors in the translation, and the coefficients of the fuzzy mapping still need to be optimized. Thirdly, since the algorithm used cannot be analyzed well in context, the articulation of the phrases is not perfect, which affects the overall effect of the translation. However, these shortcomings can be alleviated by further improving the translation memory and the algorithm.

6 Summary and Prospect

The traditional machine translation system can hardly select the correct translation according to the context, resulting in poor semantic coherence of the translation, which can not meet the user's needs well.

To address a series of problems of traditional machine translation, this paper designs an English translation proofreading system by constructing a semantic ontology translation model, building a translation memory, performing fuzzy mapping on the input utterance, selecting the correct translation according to the decision function, and outputting the final translation after calibration. The system consists of five modules. Users are able to use the search module to search the meaning of terms, and the backend engineer can analyze the behavior data of the system through the behavior log to solve the problems of the system and make improvements.

The experimental verification indicates that the English translation proofreading system designed in this paper can proofread the wrong parts of English translation process, improve the accuracy of English translation results, and compare with similar systems, not only the accuracy of translation is high, but also the translation is smooth and coherent. This system improves the efficiency and accuracy of English proofreading, reduces the cost of manual proofreading, and can meet the users' needs for proofreading English translation.

With the emergence and development of the Internet, the task of studying machine translation will become more urgent. It is noteworthy that the translation quality of machine translation system today cannot fully meet people's needs, the accuracy and coherence of translation still need to be improved, and the quality of translation has not yet made a substantial breakthrough. There is still a long way to go before machine translation can truly meet people's needs. What is realized at present is only the English-Chinese machine translation proofreading system, however, the method and

technology it adopts is an attempt to automatically acquire practical knowledge from the real language data and adapt itself to the real language data by the new alternative method needed for the knowledge uptake of machine translation. Therefore, its application prospect will be extensive and has long-term research value and significance.

The English translation proofreading system implemented in this paper has many shortcomings and needs further improvement.

An important feature of machine translation is that the overall quality of translation will be significantly improved as the translation memory increases, and thus how to maximize the translation memory from a variety of resources will be an important issue. However, the meaning of language is constantly enriched and evolves with time, and new words are constantly generated in people's daily use, and old words are constantly generated with new meanings, so timely updating of translation memory is also an important issue.

The management of the translation memory is another important issue. Because an effective translation memory must be able to handle large-scale searches at a fast enough speed, but when the size of the translation memory is too large, the improvement of the search speed is limited to the improvement of the overall translation efficiency, and the establishment of sub-banks in the translation memory for the management of terms by categories is a solution.

## Data Availability Statement

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

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

## References

Balanos-Medina, A. 2012. The Key Role of the Translation of Clinical Trial Protocols in the University Training of Medical Translators. *Journal of Specialised Translation* 17:17–36.
Bo, W. 2020. Research on Technologies of Evaluation and Diagnosis of Machine Translation System. Dissertation for the Doctoral Degree in Engineering.

Boonchieng, W., J. Chaiwan, B. Shrestha, M. Shrestha, A. J. O. Dede, and E. Boonchieng. 2021. mHealth Technology Translation in a Limited Resources Community—Process, Challenges, and Lessons Learned from a Limited Resources Community of Chiang Mai Province, Thailand. *IEEE Journal of Translational Engineering in Health and Medicine* 9:3700108. doi:10.1109/JTEHM.2021.3055069.

Brightman, A. O., R. L. Coffee Jr, K. Garcia, A. E. Lottes, T. G. Sors, S. M. Moe, and G. R. Wodicka. 2021. Advancing medical technology innovation and clinical translation via a model of industry-enabled technical and educational support: Indiana Clinical and Translational Sciences Institute's Medical Technology Advance Program – ERRATUM. *Journal of Clinical and Translational Science* 5 (1):e131. doi:10.1017/cts.2021.799.

de Boer, A., G. Villa, O. Bane, M. Bock, E. F. Cox, I. A. Dekkers, P. Eckerbom, M. A. Fernández-seara, S. T. Francis, B. Haddock, et al. 2022. Editorial for "Consensus-Based Technical Recommendations for Clinical Translation of Renal Phase Contrast MRI". *Journal of Magnetic Resonance Imaging* 55 (2):323–35. doi:10.1002/jmri.27419.

Deng, T. 2021. "The Production-Oriented Approach to Teaching Translation Between English and Chinese in Vocational and Technical Colleges", *Proceedings of the 7th International Conference on Social Science and Higher Education (ICSSHE 2021)*, Advances in Social Science, Education and Humanities Research, volume 598.

Dirnagl, U., G. N. Duda, D. W. Grainger, P. Reinke, and R. Roubenoff. 2022. Reproducibility, relevance and reliability as barriers to efficient and credible biomedical technology translation. *Advanced Drug Delivery Reviews* 182:114118. doi:10.1016/j.addr.2022.114118.

Doherty, C. L., and B. T. Buckley. 2021. Translating Analytical Techniques in Geochemistry to Environmental Health. *Molecules* 26 (9):2821. doi:10.3390/molecules26092821.

Fan, H. 2021. Application of Computer Aided Translation in Technical English Manual. *Journal of Physics: Conference Series*, 1961 (1):012041. IOP Publishing. doi:10.1088/1742-6596/1961/1/012041.

Feng, Y., W. Xie, S. Gu, C. Shao, W. Zhang, Z. Yang, and D. Yu. 2020. "Modeling Fluency and Faithfulness for Diverse Neural Machine Translation". *The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*.

Galan-Manas, A. 2011. Manas. Translating Authentic Technical Documents in Specialised Translation Classes. *Journal of Specialised Translation* 16:109–25.

He, W., Y. Wu, and X. Li. 2021. Attention Mechanism for Neural Machine Translation: A survey. *2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Xi'an, China, 1485–89. doi:10.1109/ITNEC52019.2021.9586824.

Ibragimovich, N. U. 2021. Translation Hypothesis Specifics and Problems of Translation of Scientific and Technical Multi Component Terms. *ACADEMICIA: An International Multidisciplinary Research Journal* 11 (4):1359–63. doi:10.5958/2249-7137.2021.01242.8.

Jimenez-Crespo, M. A. 2012. "Loss" or "Lost" in Translation: A Contrastive Genre Study of Original and Localised Non-Profit US Websites. *Journal of Specialised Translation* 17:136–63.

John-Jr, B. B. 2012. *Scientific and Technical Translation Explained: A Nuts-and-Bolts Guide for Beginners*, 229. Oxon & New York: Routledge

Krüger, R. 2016. The Textual Degree of Technicality as a Potential Factor Influencing the Occurrence of Explicitation in Scientific and Technical Translation. *Journal of Specialised Translation* 26:96–115.

Li, H. 2020. "Technical English Translation Techniques and Strategies", *2nd 2020 International Conference on Humanities, Cultures, Arts and Design (ICHCAD 2020) 434-437*. Australia Sydney.

Makhmudjonovna, P. M. 2021. Theory and Problems of Russian Language Translation in the Field of Technology. *ACADEMICIA: An International Multidisciplinary Research Journal* 11 (4):1173–77. doi:10.5958/2249-7137.2021.01240.4.

Mamnunah, M. A., and N. Mamnunha. 2020. Review of Altarabin (2020): The Routledge Course on Media, Legal and Technical Translation: English-Arabic-English. *Translation and Translanguaging in Multilingual Contexts* 8 (1):96–100. doi:10.1075/ttmc.00084.mam.

Mansor, I. 2021. Explicitation in the Intercultural Communication of Technical Culture in Arabic-Malay Translation of Rihlat Ibn Battuta. *Journal of Intercultural Communication Research* 50 (6):556–70. doi:10.1080/17475759.2021.1884491.

Maylath, B., and K. S. Amant, Eds. 2019. *Translation and Localization: A Guide for Technical and Professional Communicators*. Routledge: Taylor and Francis Group.

Miana, L. A., M. Nathan, D. F. Tenório, V. Manuel, G. Guerreiro, N. Fernandes, C. V. D. Campos, P. V. Gaiolla, R. Sá Cassar, A. Turquetto, et al. 2021. Translation and Validation of the Boston Technical Performance Score in a Developing Country. *Brazilian Journal of Cardiovascular Surgery*. 36(5):589–98. doi:10.21470/1678-9741-2021-0485.

Nazzal, A. R., and M. F. Khmous. 2021. Inaccuracy in the Translation of Dentistry Terms from English into Arabic. *Studies in Linguistics and Literature* 5 (1):10–27. doi:10.22158/sll.v5n1p10.

Ren, W., B. Ji, Y. Guan, L. Cao, and R. Ni. 2022. Recent Technical Advances in Accelerating the Clinical Translation of Small Animal Brain Imaging: Hybrid Imaging. Deep Learning, Transcriptomics. *Frontiers in Medicine* 9:771982. doi:10.3389/fmed.2022.771982.

Tauchi, A. 2020. Sterilization Technology Using an Ultraviolet-Radiation Source: Translated Papers. *Journal of Science and Technology in Lighting* 44:12–13. doi:10.2150/jstl.IEIJ20A000007.

Vandepitte, S. B. M., B. Mousten, S. Isohella, and P. Minacori. 2016. Multilateral Collaboration between Technical Communicators and Translators: A Case Study of New Technologies and Processes. *Journal of Specialised Translation* 26:3–19.

Wang, Z. 2021. An Empirical Study of Online Terminology Translation Retrieval: Using Search Engines as a Tool. *Journal of Physics: Conference Series* 1966 (1):012029. doi:10.1088/1742-6596/1966/1/012029.

Wook-Dong, K. 2021. Translation and Textual Criticism: Typographical Mistakes in Modern Korean Poems. *Acta Koreana* 24 (2):55–74. doi:10.18399/acta.2021.24.2.003.

Wright, L. D., and S. E. Wright, eds. 1993. *Scientific and Technical Translation*. American Translators Association Scholarly Monograph Series VI:Amsterdam

Zhu, P. 2021. Audience and Purpose as a Guide to Improve Acceptability and Readability of Technical Translation. *International Journal of Linguistics, Literature and Translation* 4 (7):98–108. doi:10.32996/ijllt.2021.4.7.11.