

Research Article

Innovative System for Analyzing English Translation Software Based on IoT Big Data

Rongliang Niu^{1,2} and Wei Yang ¹

¹*School of Foreign Languages, Zhanjiang University of Science and Technology, Zhanjiang 524088, Guangdong, China*

²*School of Humanities, Universiti Sains Malaysia, Penang 11800, Malaysia*

Correspondence should be addressed to Wei Yang; yangwei@zjkju.edu.cn

Received 9 April 2022; Revised 5 May 2022; Accepted 19 May 2022; Published 14 June 2022

Academic Editor: Yanyi Rao

Copyright © 2022 Rongliang Niu and Wei Yang. 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 expansion and deepening of international exchanges and cooperation, language differences have become the biggest obstacle to exchanges and cooperation. At present, English translation software is the translation between natural languages made by humans on the basis of computers. As an important technical means to break through language barriers, machine translation is a software that uses computers to process spoken or written language. This paper aims to study the innovative system of English translation software, using the Internet of Things and big data algorithms to analyze some loopholes and deficiencies in the innovative system, and provide some data research for the system update. This paper proposes to analyze the innovative system of translation software based on the big data of the Internet of Things (IoT). Among them, the big data of the Internet of Things chooses the dynamic frame time slot ALOHA algorithm and the lower limit estimation algorithm. Its application in system analysis is very innovative. The system consists of a translation engine and a machine dictionary, which solves the basic problem of insufficient translation database of current translations, and plays a role in helping the English translation software innovation system to improve the accuracy of translations. The experimental results of this paper show that the translation system's translation software achieves 61.3% of the A-target fluent translations, and 22.7% of the B-targets can translate the original text. It can be seen that English translation software has important research significance.

1. Introduction

With the increasing globalization of the global economy and the continuous and rapid development of the world's Internet applications, international Internet exchange activities and scientific and technological cooperation will become more and more frequent and the division of labor is meticulous. Language differences have become the two biggest language barriers in the current international information technology exchange and learning and scientific and technological cooperation and development. It is also becoming increasingly important in the global economic development environment and the field of contemporary social life, and it is one of the main focuses of current research and discussion.

Entering the 21st century, with the gradual intensification of international cooperation, exchanges and competition in society, and the arrival of the global Internet era, the

level of social informatization in society as a whole will inevitably become higher and higher. In order not to be gradually eliminated by the tide of this era, people have to keep up with the tide of the times and always pay attention to the latest social dynamic information at the forefront. Most of the above dynamic information materials are also the first time that they appear publicly in the form of pure foreign languages on the modern Internet media. Therefore, despite the rapid development of Internet language in recent decades, the existing language machine translation system is still unable to guarantee a satisfactory overall translation effect, especially the translation of some complex sentence sequences and complex chapters. Therefore, the comparative research method of machine language translation effect should not only have important scientific value, but also must have practical application value. The most meaningful point of view is that under the requirements of the

development situation in the new century, it is of extremely practical and far-reaching significance to further develop and study language machine semantic translation comparison technology and improve the overall translation effect in the existing machine language translation comparison system.

The novelty of the article is as follows: (1) this article analyzes the English translation software system by using the method of Internet of Things big data, which greatly improves the analysis efficiency and the evaluation speed of system. (2) Based on the theoretical basis of designing a language framework system, this paper successfully designs and implements other practical functions in the machine vision translation language system. One of the key contents is the realization of the grammar design function of the English-Chinese electronic dictionary system. And it realizes the automatic alignment of English-Chinese sentence structure and the semantic design function of English-Chinese sentence translation system. (3) The translation thesaurus system specially researched and manufactured based on the big data platform of the Internet of Things can also be easily used for various expansion methods such as increasing the number of translation instances and expanding the translation vocabulary. Therefore, it can help people avoid some traditional rules-based machine language translation comprehension methods and apply this method to some more deep linguistic problems and analysis that must be reinterpreted.

2. Related Work

The Internet of Things is the “Internet of Everything Connected,” which is an extension and expansion of the Internet based on the Internet. It is a huge network formed by combining various information sensing devices with the network, realizing the interconnection of people, machines, and things at any time and any place. Tallapragada V tracked customer sentiment and provides customer behavioral insights through an IoT-integrated data intelligence system running on Apache Spark Cluster [1]. Philemon Kibiwott proposed a more efficient scheme that borrows computing power from cloud servers to handle expensive computations, while leaving simple operations to users [2]. Jamil et al. proposed a new approach in which IoT data are processed and analyzed in real time using big data tools [3]. Ahmad and Afsal believed that fog computing originates from cloud computing and is implemented on the end user’s device to deal with network latency [4]. Shin et al. applied a system for processing informal data by collecting, storing, processing, and analyzing data [5]. Rajeswari et al. used IoT devices to sense agricultural data and store it in a cloud database, using cloud-based big data analytics to analyze the data [6]. Zhao et al. proposed an IoT data credibility detection method based on regional context, which can effectively detect point anomalies, behavior anomalies, and context anomalies [7]. Granat et al. proposed an event detection method for one-dimensional data streams, which relies on an event intensity function and is an extension of the typical “true or false” decision scheme [8]. Wei et al. provided a comprehensive

review of IoT data flow management. We first analyze the key challenges faced by IoT data flow and provide a preliminary overview of related technologies in data flow management, spanning data flow perception, mining, control, security, privacy protection, etc. [9]. Palaniswami et al. presented a useful set of visual assessment tools and techniques for clustering trend (VAT). It further highlights how these technologies are advancing IoT through large-scale IoT implementation [10]. Meerja et al. studied IoT data algorithms to share their data with the online world to obtain global knowledge and information with high commercial value [11]. This work by Almeida A involves key performance analysis of an IoT-aware ambient-assisted living (AAL) system for elderly monitoring. The analysis focuses on three main system components: (i) a city-wide data acquisition layer, (ii) a centralized cloud-based data management repository, and (iii) a risk analysis and forecasting module [12]. Ko and Kim’s goal in research is to develop a healthcare platform. The platform can receive diabetes information based on remote input, storage, analysis, processing, and visualization of diabetes information generated by various IoT [13]. However, the practicality of these experiments is not too strong, and they only stay in the preliminary theoretical stage.

3. IoT Big Data

3.1. Dynamic Frame Slot ALOHA Algorithm. When the dynamic frame slot ALOHA algorithm faces the number of tags existing in the tag reader, the number of dynamic frame slot tags existing in a dynamic frame domain can also be changed. It is one of the most widely used DFSA algorithms in the Internet of Things [8]. In the DFSA algorithm, it is assumed that the initial frame rate range of the reader to be recognized is L , and the range of the number of tags within the time range set by the reader to be identified is m . Then, the probability that the reader is set with t number of tags at the same time slot to be selected and identified at the same time is

$$P_{m,1/L} = C_m^L * \left(\frac{1}{L}\right)^t * \left(1 - \frac{1}{L}\right)^{m-t}. \quad (1)$$

According to formula (1), the expected value of the idle time slot when $t=0$ can be obtained:

$$\begin{aligned} E_0 &= L * P_{m,1/L(0)}, \\ &= L * \left(1 - \frac{1}{L}\right)^m. \end{aligned} \quad (2)$$

When $t=1$, the expected value of correctly identifying the time slot is

$$\begin{aligned} E_1 &= L * P_{m,1/L(1)}, \\ &= m * \left(1 - \frac{1}{L}\right)^{m-1}. \end{aligned} \quad (3)$$

The expected value of the collision slot when $t \geq 2$ is

$$E_t = L - E_0 - E_l. \quad (4)$$

After one round of recognition process, the correct recognition rate of the whole system is

$$\begin{aligned} R &= \frac{E}{L}, \\ &= \frac{m}{L} * \left(1 - \frac{1}{L}\right)^{m-1}. \end{aligned} \quad (5)$$

According to formula R , the derivation of the frame length L can be obtained:

$$\frac{dR}{dL} = \frac{d}{dL} \left(\frac{m}{L} * \left(1 - \frac{1}{L}\right)^{m-1} \right) = 0. \quad (6)$$

It can be obtained from the above formula that when $L \approx m + 1$, the working state of the system reaches the best, and theoretically the maximum throughput rate of the system can be obtained from the following formula to be 36.8%.

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{m}{L} * \left(1 - \frac{1}{L}\right)^{m-1} \\ = \lim_{m \rightarrow \infty} \left(1 - \frac{1}{m}\right)^{m-1} \approx 0.36788. \end{aligned} \quad (7)$$

Through the logical derivation and analysis of the above formula, it can be known that during each round of tag recognition in the system, the DFSA algorithm dynamically adjusts the frame length so that the frame length set by the tag reader is approximately matched or basically equal to the number of tags before the system is fully recognized by the tag. Only then can the system maximize the label throughput rate of the entire system.

3.1.1. Lower Limit Estimation Algorithm. The lower limit estimation algorithm assumes that in the identification process of each round, if there is a time slot collision phenomenon in the system, there are only two tags in each collision time slot. Therefore, in this round of label recognition process, the number of labels to be recognized in the next round of recognition process estimated by the lower limit estimation algorithm is

$$N_{\text{est}} = 2 * E_t. \quad (8)$$

The algorithm assumes that only two tags will respond if they choose the same time slot. However, in practical applications, there may be more tags to select the same time slot, so the estimated number of tags in the above situation may be quite different from the actual number of tags. It can be utilized in the instance part and the structure construction part of the system analysis. The use of dynamic frame time slot algorithm to carry out the evaluation stage of the system can make the system evaluation result more accurate.

3.1.2. Three-Dimensional Estimation Algorithm. The 3D estimation algorithm is an effective image denoising algorithm. By matching with adjacent image blocks, several similar blocks are integrated into a three-dimensional matrix, filtered in three-dimensional space, and the result is inversely transformed and fused to two-dimensional to form a denoised image. The denoising effect of this algorithm is remarkable, and the highest peak signal-to-noise ratio can be obtained, but the time complexity is relatively high.

The three-dimensional estimation algorithm is that in the process of reader identification, if t tags select time slot j at the same time, the distribution function of the time slot can be obtained according to the above formula as follows:

$$p_t \{ \text{Collision occurs when } t \text{ labels simultaneously select time slot } j \} = \begin{cases} 0 & t = 0, 1, \\ \frac{p_t}{1 - p_0 - p_1} & t \geq 2. \end{cases} \quad (9)$$

Then, the average number of tags in slot j is

$$e_t = \sum_{t=2}^l t * p_t = \frac{\sum_{t=2}^l t * C_m^t * (L-1)^{m-1}}{L^m - (L-1)^{m-1} * (L-1+m)}. \quad (10)$$

The three-dimensional estimation algorithm estimates the number of tags according to the number of tags e_t and collision probability P_c in the collision time slot in the system, and its expression is as follows:

$$\begin{aligned} e_t &= 1.5 * P_c + 2 P_c < 0.6, \\ e_t &= 12^{P_c-0.6} + 1.90.6 \leq P_c \leq 0.8, \\ e_t &= 6^{3*(P_c-0.6)} + 2.53 P_c > 0.8. \end{aligned} \quad (11)$$

The number of tags to be identified in the next round is

$$U_{\text{est}} = e_t * E_t. \quad (12)$$

3.1.3. Schoute Estimation Algorithm. In the Schoute estimation algorithm, assuming that the number of tags in the selected time slot j obeys the Poisson distribution with a mean of 1, then the probability that time slot j is simultaneously selected by t tags is

$$P_t \{ t \text{ tags simultaneously select time slot } j \} = e^{-1} * \frac{1}{t!}. \quad (13)$$

When the time slot collides, the number of tags in the selected time slot j can be obtained by the formula:

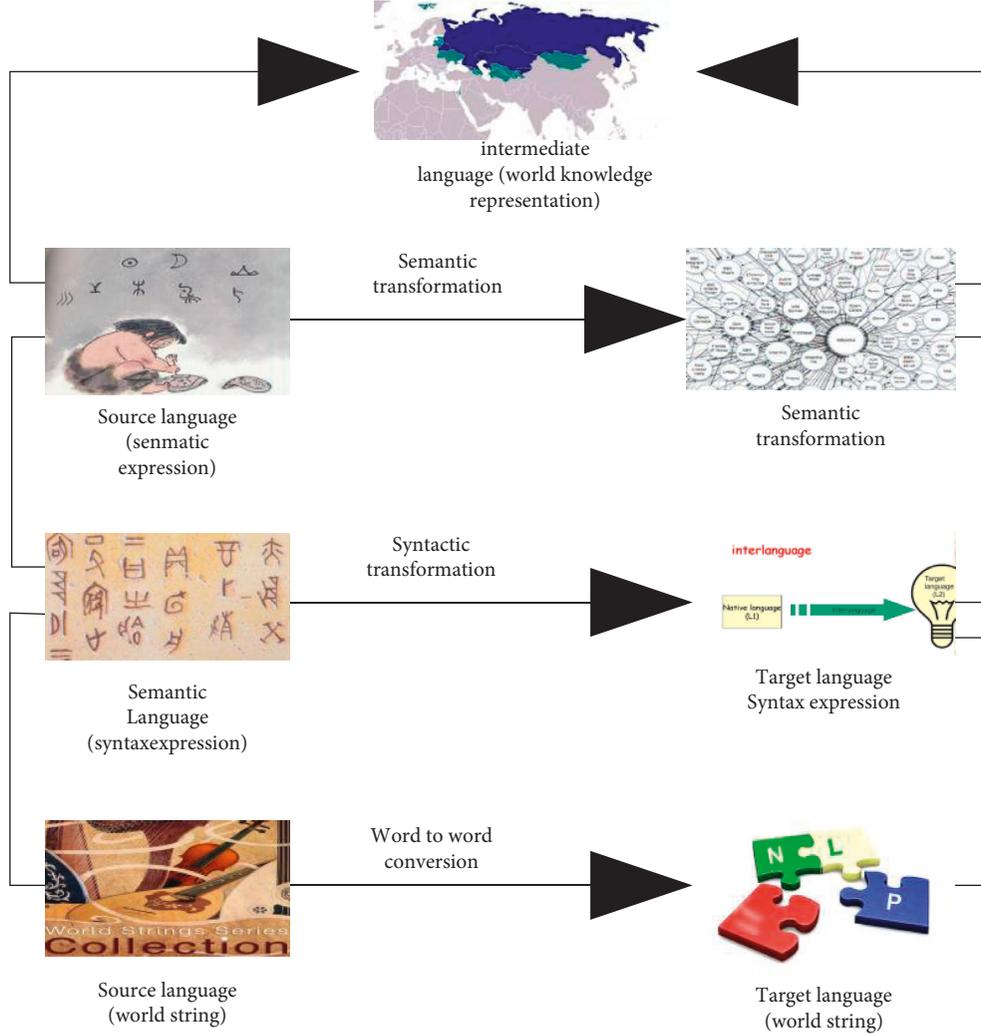


FIGURE 1: Different rule-based machine translations.

$$\sum_{t=2}^{\infty} t * p_t = \frac{\sum_{t=2}^{\infty} e^{-1} * 1/(t-1)!}{1 - 2e^{-1}} \quad (14)$$

$$= \frac{e-1}{e-2} \approx 2.39.$$

Then, the number of tags to be recognized is

$$\text{Nest} = 2.39 * E_t. \quad (15)$$

3.2. Ant Colony Algorithm in Big Data. Assuming that in a certain environment network, there are m routing request points waiting for routing processing, there are at least n routing access points waiting for routing to perform tasks, and there are at least x ants in each routing group network. Ants will only randomly leave a small amount of pheromone on a certain path node during the whole process of foraging. The pheromone concentration will gradually volatilize with the arrival time. The next ant will only randomly re-select the path based on the remaining pheromone concentration after reaching that moment [9]. The probability that the k th ant

chooses to transition from path i to path j is P_{ij}^k , and then the calculation formula of P_{ij}^k is as follows:

$$P_{ij}^k = \begin{cases} \frac{(\tau_{ij}^{\alpha})}{\sum_{b \in \text{allow}_k} (\tau_{ij}^{\alpha})(\eta_{ij})^{\beta}}, & j \in \text{allow}_k \\ \text{else} & \end{cases} \quad (16)$$

Among them, τ_{ij} represents the pheromone in the set from path i to path j , η_{ij} represents the path length from path i to path j , allow_k represents the set of paths that the k th ant is allowed to pass through (unvisited), α and β are control parameters, and the value range is $[0, 1]$.

When ants are looking for the optimal path, the update strategy of pheromone concentration is as follows:

$$\tau_{ij}(t+1) = (1 + \theta)\tau_{ij} + \sum_x \Delta\tau_{ij}^k(t),$$

$$\Delta\tau_{ij}^k(t) = \begin{cases} \frac{Q}{d_k}, & \in \text{allow}_k, \\ 0, & \text{else}. \end{cases} \quad (17)$$

It can be seen from the above that the use of dynamic frame time slot ALOHA algorithm, lower bound estimation algorithm, and three-dimensional estimation algorithm in the Internet of Things and big data can be used to analyze innovative systems in English translation software. Then, there is a clearer analysis of its structure, rules, and some specific instances, and the ant colony algorithm is used to evaluate the system, which can make the evaluation results more accurate.

4. English Translation Software

4.1. English-Chinese Machine Translation System

4.1.1. Rule-Based Approach. The language method based on the study of language rules assumes the logical process of translation itself to understand and analyze other kinds of languages [14, 15]. Whether this is a transition language-based translation method or an intermediate transition language-based translation, the emphasis is only on the difference between their comprehensive analysis ability and in-depth mastery of the connotation problems of the switch language and the degree of high abstraction and generalization of the knowledge of the switch language and the connotation of the expression, as shown in Figure 1.

Rule-based methods have always dominated the machine translation community and still play an important role today, and the influential MTSs are now rule-based. However, compared with the traditional rule-based method, the current rule-based method has undergone many changes. These changes are mainly reflected in the following: in the acquisition of rules, traditional rule methods mainly rely on linguists to summarize rules and debug them. However, more emphasis is now placed on automatically obtaining rules from corpora. Traditional rule methods tend to focus on describing coarse-grained and globalized knowledge of large-scale linguistic rules. However, more attention is now paid to describing fine-grained, local, and small-scale linguistic knowledge, showing a trend of “small rule base, big dictionary.”

To really develop such a machine semantic translation system based on natural language rules, people must first design a semantic knowledge representation system. All the language knowledge that may need to be used in the process of natural language translation is expressed in the form of an operational language that can be implemented on a computer [16–18], as shown in Figure 2.

4.1.2. The Overall Frame Structure of the System. The system framework is roughly divided into the following three parts:

(1) Translation System Knowledge Base. The learning process of machine language translation can actually be regarded as a process of learning language application-related knowledge to perform semantic reasoning operations. The representation of knowledge is the theoretical basis for understanding the translation process [17]. People often divide the various kinds of knowledge used in machine intelligent translation language into two categories according to the

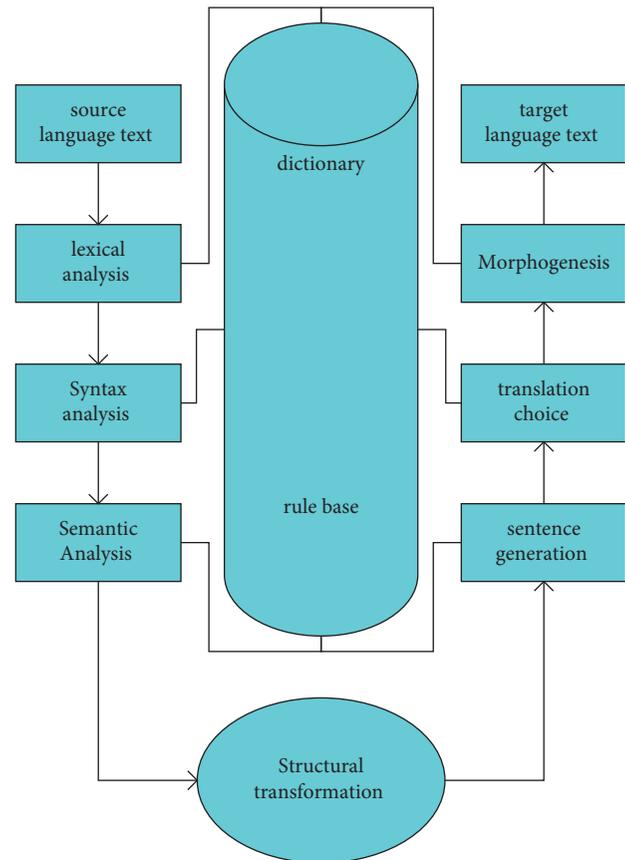


FIGURE 2: Rule-based transformation.

form of representation, such as knowledge inside the machine and knowledge outside the machine.

(2) System Processing Part. The English vocabulary processing part includes automatic English word segmentation processing and English word segmentation processing. These two parts mainly include the theoretical basis of phrases and sentence combinations of English sentence patterns. The automatic word segmentation system uses the highest matching algorithm. And the system can use various rule analysis and statistical analysis methods. In the part of automatic elimination and integration of English rule analysis, the system is divided into two parts: phrase analysis and sentence pattern matching.

(3) User Interface. The debug user interface used is usually divided into a system client and a debug user interface that can be managed by a system administrator. A very good visual effect and practical and concise machine learning translation and dictionary system compiler compilation management page and the management page of the entire compiler debugging work system, which can be specially designed to use a machine translation system, and the operator can also be very intuitive and can simply and accurately complete the compilation and management of all language rules. Therefore, the task implementation efficiency in the entire compiler and debugging work system can be greatly improved, and the overall work efficiency of the

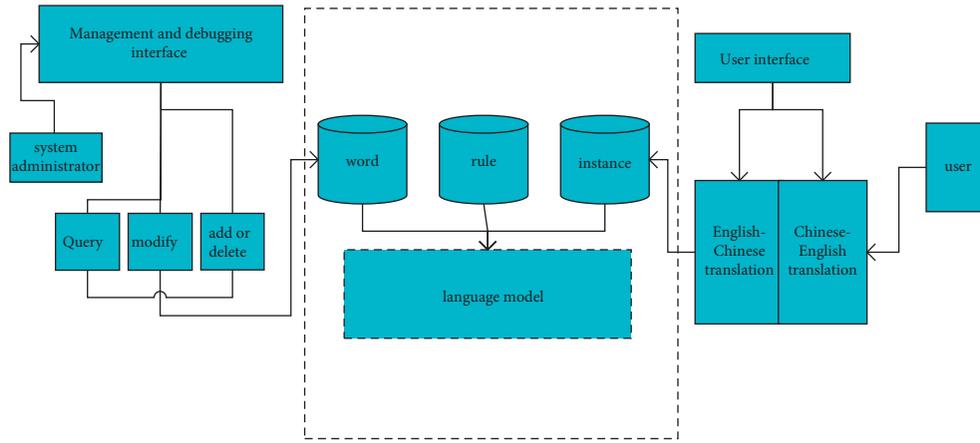


FIGURE 3: Overall architecture diagram of the system.

entire language knowledge base work system can be improved more efficiently, quickly, and accurately. The overall frame structure of the system is shown in Figure 3.

The knowledge management and debugging function interface that comes with the application system is responsible for two parts: knowledge base establishment, maintenance, operation, and knowledge translation and debugging management of natural language processing [18]. The system administrator can directly display the process of generating any syntactic component of the system and other syntactic feature nodes (attributes and values) corresponding to the system according to their needs.

4.1.3. System Translation Algorithm and Workflow. Many times, when people learn a new language, they learn it the way they did before. In many cases, it is often necessary to solve new grammatical problems based on the experience of people using some translation from the original language [19]. The main language translation information screening workflow involved in this translation system is roughly shown in Figure 4.

Among them, the rule method mainly refers to the translation design method based on transformation, which follows the translation design principle of generating rules while analyzing [20]. The specific translation algorithm is shown in Figure 5.

The input of the original text can use text files, keyboard input, and scan input:

If the text does not have any disambiguation rules, the text is automatically defaulted to the first part of speech. The improved figure parsing algorithm is automatically adopted in the stage of structural analysis. Most of the structural transformation steps use a local subtree transformation algorithm based on a combination of top-down and bottom-up [21]. In the process of constructing the structure, bottom-to-bottom local subtree transformation algorithms are applied, respectively.

4.1.4. Part-of-Speech Tagging and Structural Tagging of Phrases. The part-of-speech process is a scientific process in

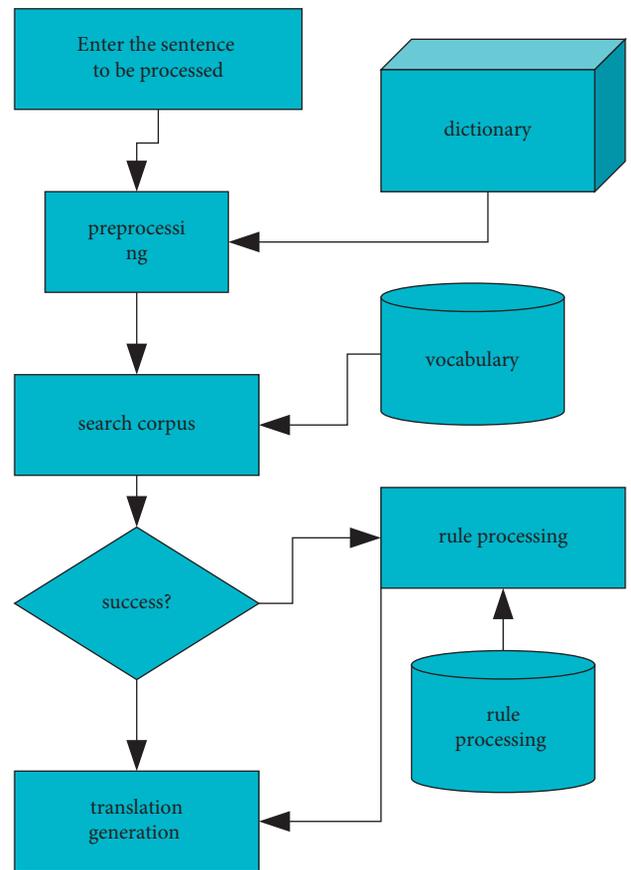


FIGURE 4: English-Chinese translation process.

which people comprehensively analyze and classify the basic part-of-speech system composed of various words involved in the entire sentence structure and standardize their use. There are mainly open parts of speech, which can continuously appear in large numbers of new English words. The four main types of open language parts of speech that appear in the contemporary English-speaking world are shown in Table 1.

Secondary part of speech is also called closed part of speech. Generally speaking, the number of words in these

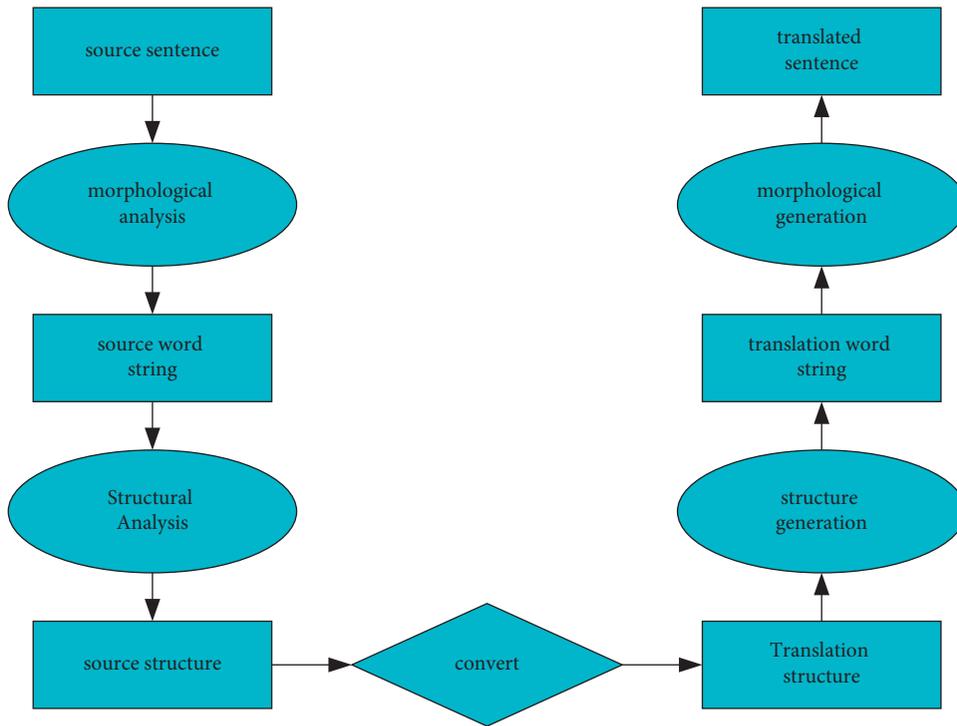


FIGURE 5: The translation process for rule processing.

TABLE 1: Main parts of speech in English.

Noun:	Student	Linguistics	Lecture
Verb:	Like	Read	Went
Adjective:	Tall	Lovely	Red
Adverb:	Loudly	Constantly	Hard

parts of speech remains fixed, and generally no other new words are allowed to be added. There are mainly six types of closed parts of speech in modern English vocabulary, as shown in Table 2.

Several words of different types of parts of speech can be individually combined into the same English phrase form or phrase. For example, in the process of syntactic structure analysis in English, one or several English words can form the following phrase structure in Table 3.

In the process of annotating English sentences, different sets of annotations are usually used in different corpora. In order to make the sentence structure more unique and recognizable, this system does not use standard English part-of-speech tagging rules and tagging sets based on context-free grammar. The annotation sets used in this system are shown in Table 4.

The conjugation forms of verbs in English are more complicated. For example, although the infinitive is a form of the verb, it can also function as a noun and retain some characteristics of the verb, so it is also marked as the infinitive phrase inf alone. The gerund form is also a form of the verb, which is also marked separately with the ger tag [22]. In addition, the past participle form of the verb is also marked with ppl.

TABLE 2: Secondary parts of speech in English.

Determiner	The, a, an
Auxiliary	Can, will, do, be, have
Preposition	In, at, to, on
Pronoun	He, she, us, mine
Conjunction	And, or, but, while
Interjection	Oh, ah, ouch

TABLE 3: Secondary parts of speech in English.

Noun phrase	A red brick, the beautiful lake
Verb phrase	Sing a song, write in ink
Preposition phrase	On the table, at the gate
Adjective phrase	Very clever, quite able

TABLE 4: Annotation sets used by this system.

Np:	A red brick, the beautiful lake
Vp:	Sing a song, write in ink
Pp:	On the table, at the gate
Adjp:	Very clever, quite able
Advp:	Adverb phrase
S:	Sentence
Inf:	Infinitive
Idiom:	Idiom

In order to facilitate the alignment of English and Chinese words and phrases, the Chinese annotation set is roughly the same as the English one. However, Chinese does not have a variety of verb forms, so at this stage, the Chinese tagging set of this system is a subset of the English tagging set.

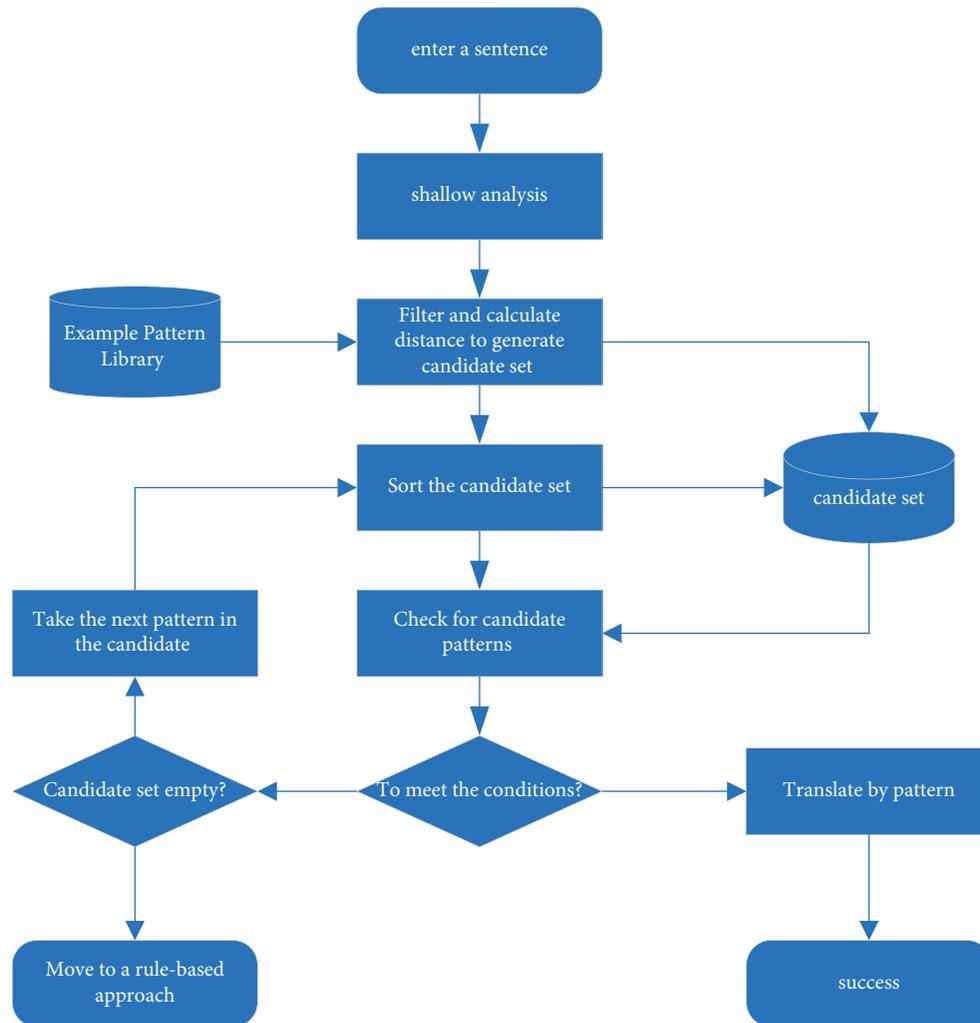


FIGURE 6: Matching algorithm flow.

4.2. Key Technologies of English-Chinese Translation System

4.2.1. Instance Pattern Matching. The strategy of instance pattern matching is to apply the input sentence and the results of lexical analysis, part-of-speech tagging, and shallow syntactic analysis of the input sentence to match the instance patterns in the library, that is, to calculate the similarity between the two [23]. Select the most basic similar instance pattern in the library as the matching result. At the same time, according to its target language pattern and phrase target pattern generation algorithm, it constructs a sentence and outputs all the translations in the sentence. The algorithm flow of instance pattern matching is shown in Figure 6.

4.2.2. Multi-Engine Translation System. Due to the limitation of corpus size, it is difficult for instance-based machine translation to achieve a high matching rate; people can use a relatively complete standardized translation method, combined with a sample library to take advantage of sequential processing [24].

In view of this, it is natural for people to think of effectively combining the two methods, which will definitely improve the quality of machine translation. As a result, the concept of multi-engine was proposed. The meaning of multi-engine is that a machine translation system adopts multiple translation methods, and each method is an engine, as shown in Figure 7.

It can be seen from the above that we can use the latest and most mature English computer software technology in the world, the advanced and perfect research results of contemporary English grammar theory, and a newly designed English-Chinese bilingual corpus. Constructing a most preliminary and complete contemporary English-Chinese machine language translation application system is the frontier topic that we focus on exploring and solving in this paper [25].

4.2.3. The Overall Design of the Dictionary Base Class. The division principle between the dictionary base class and the dictionary subclass in the dictionary class library: Various operations on the dictionary records of automatic

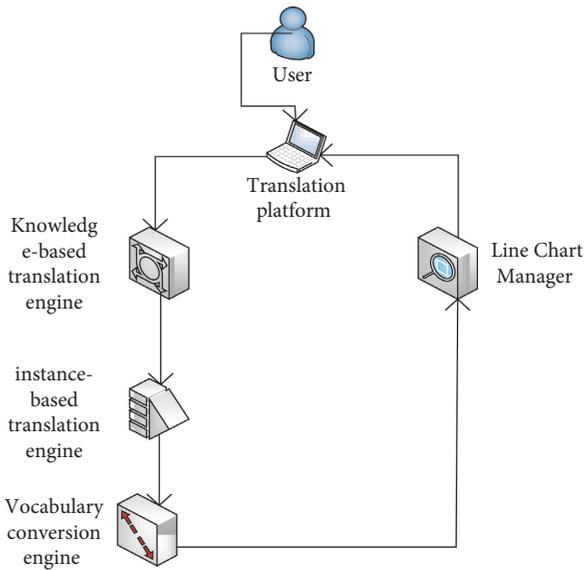


FIGURE 7: Multi-engine translation system structure.

machine translation can be simply and abstractly divided into query, add, delete, and modify operations.

From a dynamic point of view, the subclass derived from the base class dictionary class provides an external calling interface. It can be built in the interface function library model that directly calls the corresponding type of function call. Functions such as direct function calls between the function library and the corresponding subclass functions of the function are automatically converted into the function itself directly calling the functions between the corresponding base class functions [26]. Various operations on external physical files are implemented by the base class. Figure 8 depicts the dynamic model between child and parent classes.

From a static point of view, the data relationship between the subclass and the base class is to map several keywords to one keyword or map several data items to one data item. In this way, the base class can be applied to a wider range. This static relationship can be represented by Figure 9.

5. English Translation System Evaluation and Results

5.1. English Translation System Evaluation. Based on the above proposed machine translation method and analysis, matching, transformation, and object generation methods, we design and implement an English-Chinese machine translation system. The system includes knowledge base subsystem, translation subsystem, knowledge management interface, and user interface. The knowledge base subsystem and translation subsystem are the main parts of the system, which are built on the language model and translation model, respectively, as shown in Figure 10.

Language model theory refers to the research on the definition and description of basic language attributes of natural language features by human systems. It is also a systematic re-understanding of the human natural language

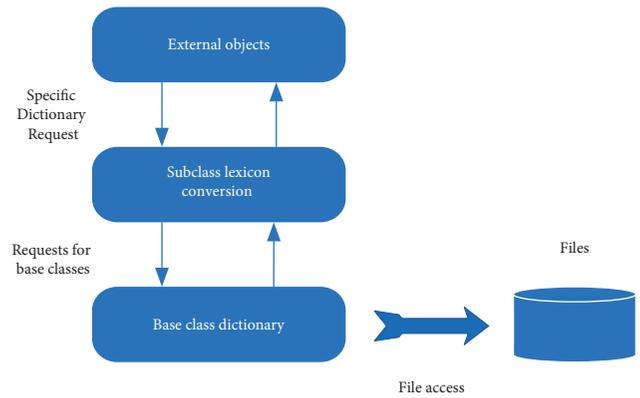


FIGURE 8: Subclass and superclass model diagram.

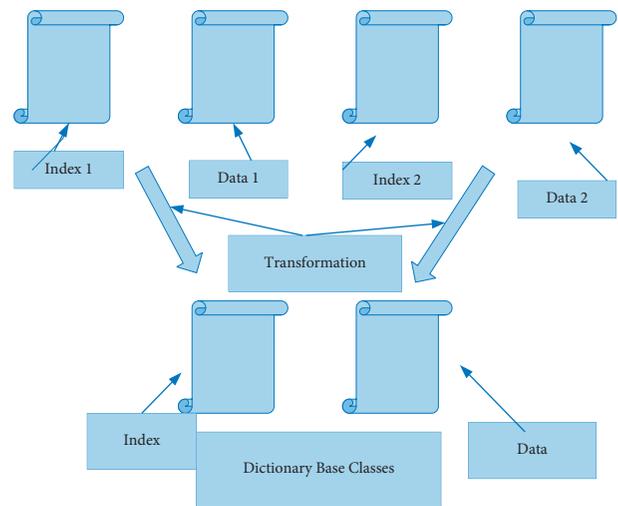


FIGURE 9: System diagram.

grammar model from the perspective of the processing law of speech information. The main types include subject parts of speech, sub-parts of speech, morphological features, semantic classes, syntax classes, sentence features, etc., in English. It is actually the mathematical basis for the application of description algorithms and translation algorithms of system language knowledge.

5.1.1. System Evaluation. In order to test the performance of the machine translation system, we organized relevant experts to form an evaluation team to evaluate the translation quality, translation speed, system maintainability, and scalability of the system. Among them, the dynamic frame time slot algorithm, three-dimensional estimation algorithm, and ant colony algorithm in the big data of the Internet of Things are used to make the results more accurate.

(1) Evaluation Method. The evaluation team consists of 1 machine translation expert, 1 linguistics expert, and 1 business unit expert. The evaluation method mainly adopts two methods: closed test and open test. The closed-ended sentence test means that each expert member of the evaluation team randomly selects some sentences from the

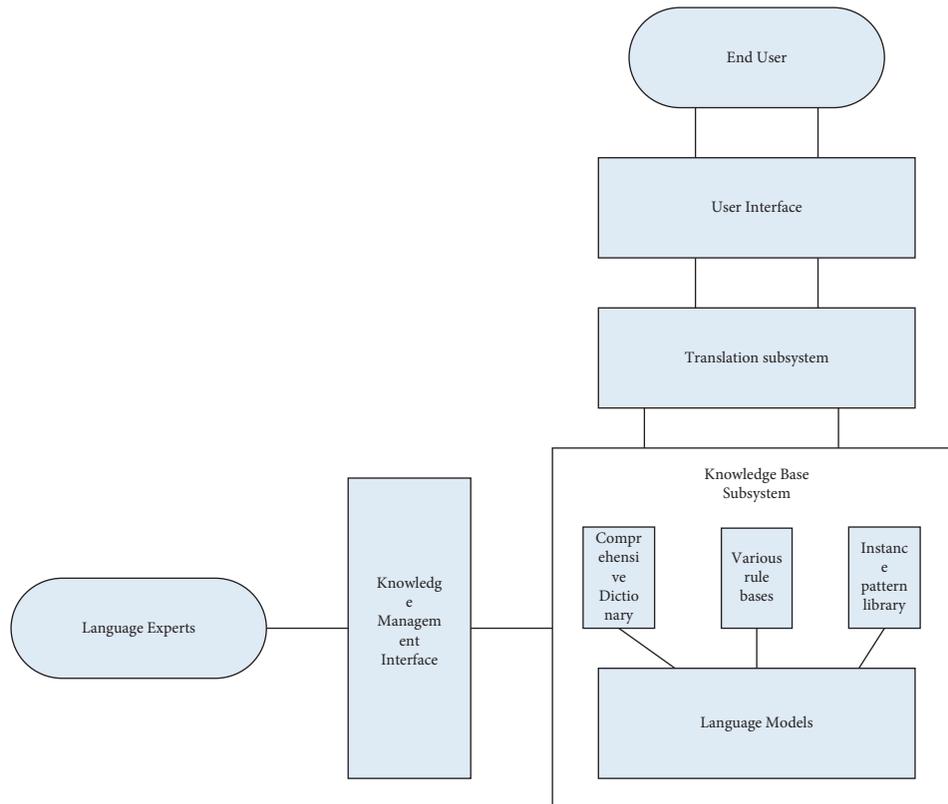


FIGURE 10: Structure of English-Chinese machine translation system.

sentences in the “English-Chinese Machine Translation Quality Test Outline” compiled by Peking University for closed-ended testing. An open-ended sentence test is a test conducted by English experts on some English sentences designed by them (due to the limited knowledge base of the system, the research group can supplement dictionary knowledge and rule knowledge on-site). After being automatically translated by the system, the experts of the evaluation team give an evaluation level to the translation quality of the translation results of each sentence according to the evaluation standards. Then, according to the evaluation method, the evaluation level is quantified and calculated, and the system translation quality evaluation value is obtained. The contribution of translation speed, system maintainability, and scalability to the system is directly scored by experts in the evaluation group. Finally, the overall evaluation value of the system is obtained, and the evaluation results are given by the evaluation team accordingly.

(2) *Evaluation Criteria.* According to the evaluation method of the machine translation system of the National 863 Plan, the translation quality is divided into 6 grades as follows: A, B, C, D, E, and F.

(A) The translation is accurate, fluent, and clear. The language accurately conveys all the various textual information contained in the translated original content in the text. Except for some small typos, there is no need to deliberately make any modifications or supplements.

- (B) The translation conveys the information of the original. People can fully understand the true meaning of the Chinese translation without referring to the original Chinese text. However, this kind of translation will inevitably have some problems in many aspects, such as grammar, choice of translated words, Chinese vocabulary expression, and customary translation, which need to be revised repeatedly.
- (C) The translation has roughly expressed the meaning of the readers’ preliminary understanding of the original content, and the partial translation content is similar to the original text or may have some discrepancies.
- (D) Part of the translation conforms to part of the meaning of the original text, but the whole sentence is not translated correctly, but the words in the whole sentence of the original text are translated, which is helpful for manual post-editing.
- (E) The translation does not make sense, or the meaning is completely wrong. However, I always feel that there are so few partial sentences or some partial words that are well translated.
- (F) All translations are not up to standard.

(3) *Level Quantification.* The quantitative evaluation of the translation quality level requires comprehensive quantification of the evaluation result level and then scoring. First, quantify the following 6 individual evaluation index grades

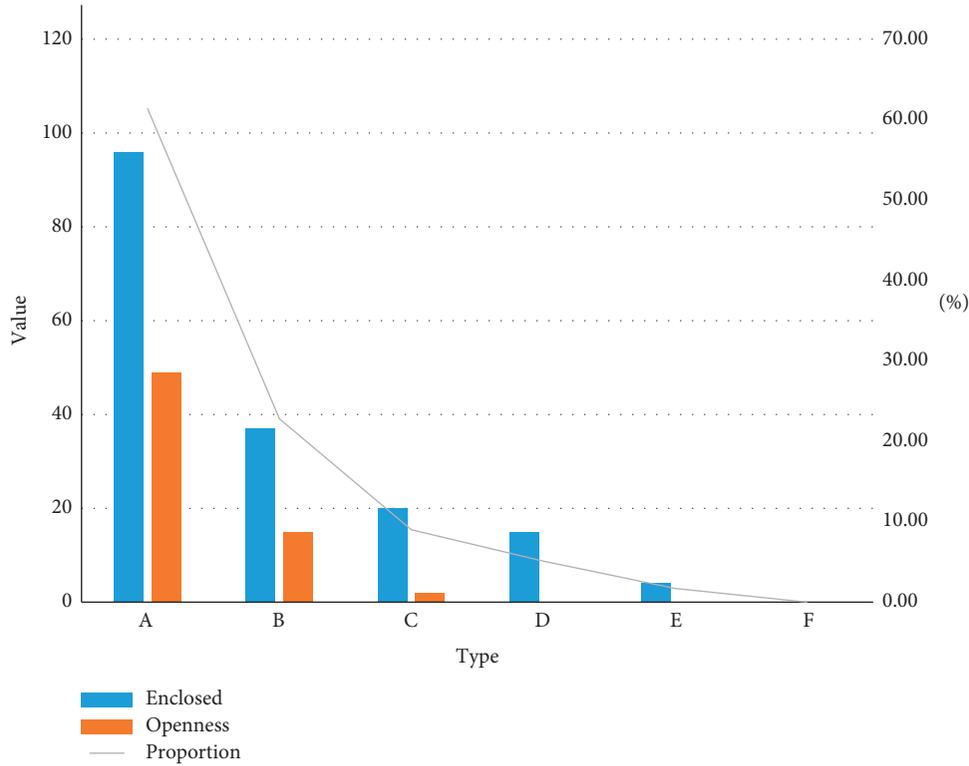


FIGURE 11: Quality of translation reviews.

in the translation quality: A-100 points, B-80 points, C-60 points, D-40 points, E-20 points, and F-0 points. Then input the corresponding Chinese-English reference sentences for each of the above indicators. After the full-text system is translated and processed, the translation expert team will give the corresponding evaluation level to the translation result according to the final translation quality score. Finally, after summarizing, the evaluation level is converted into a score, and the weighted average is calculated. For translation speed, maintainability, and scalability, the evaluation results are given by the evaluation team experts after evaluation. The total evaluation value of the system is obtained by summarizing [27].

(4) *Evaluation Results.* The closed test is to randomly select 58 simple sentences from the “Outline of English-Chinese Machine Translation Quality Test” compiled by Peking University and conduct a closed test. The open-ended test was tested by 22 simple sentences independently designed by the evaluation team. The overall test results of the three experts are shown in Figure 11.

5.1.2. *Distribution Analysis of Syntactic Equivalence Pairs.* To investigate the distribution of syntactic equivalence pairs, we count the number and length of syntactic equivalence pairs in the experiments. In Figure 12, we present the statistics on the number of syntactically equivalent pairs.

The length distribution of the syntactic structure nodes involved in the experiment is counted. The length of a syntactic node refers to the number of words and punctuation

included in the segment covered by a syntactic node in a sentence. First, we count the distribution of the number of nodes in the syntactic equivalence pairs whose length is between 1 and 50 in the syntactic structure of the reference translation. Then, we calculated the ratio of the number of nodes in syntactic equivalence pairs to all syntactic nodes of the same length at different lengths [28], as shown in Figure 13.

From the above statistical results, we can find three main information. First, in terms of absolute quantity, it is obvious that the number of small-scale syntactically equivalent nodes is much larger than that of large-scale syntactically equivalent nodes. Second, from the perspective of quantity ratio, the ratio of nodes in syntactic equivalence pairs in large-scale nodes is higher than that in small-scale nodes. This shows that among the syntactic structures of sentences with the same semantics, consistent syntactic structures are more likely to appear in the high-level structures where large-scale nodes are located, and syntactic equivalence pairs exist.

(3) *Distance Distribution Analysis of Wrong Word Order Conversion.* Next, the distribution of distance information in Bilingual Error Word Order Transformation (Order-iBT) is calculated. First, the distribution of the reference translation lexical distance (RDt) of all Order-iBTs in Chinese-English and English-Chinese translations is calculated, as shown in Figure 14.

Figure 14 gives two pieces of information. First, the number of RDts with short distances is much larger than those with long distances, where ranking errors with distances 2 and 3 are the most common. Second, compared with Chinese-English translation, the distance of target

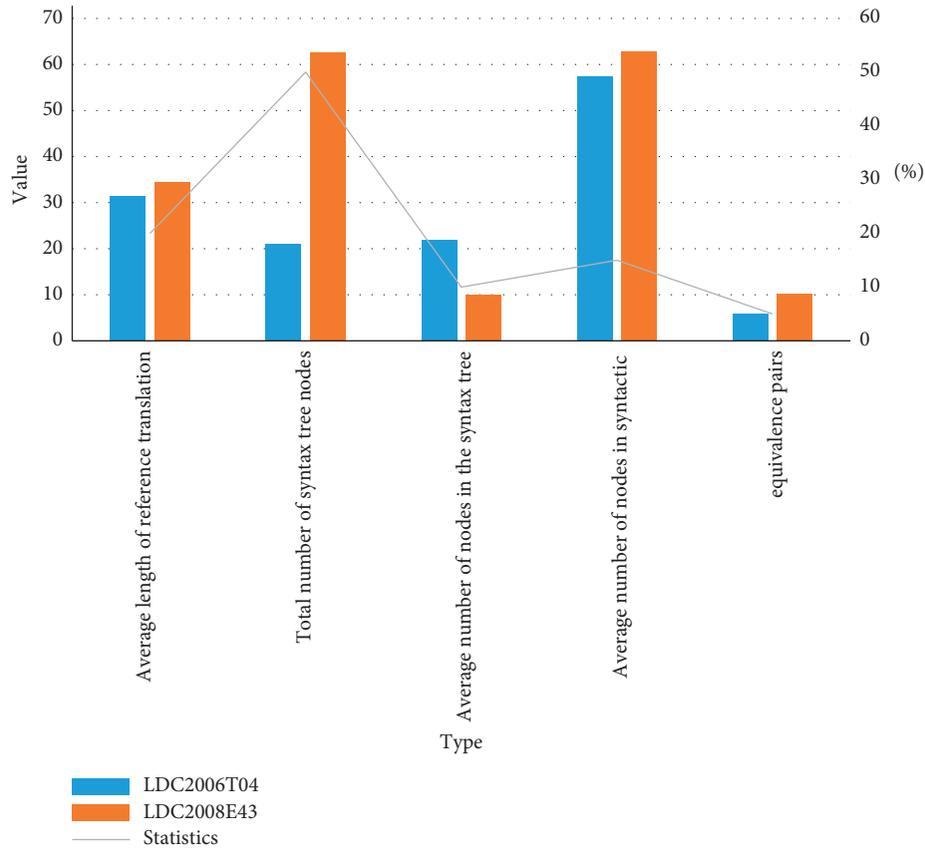


FIGURE 12: Node statistics in the reference translation syntax tree.

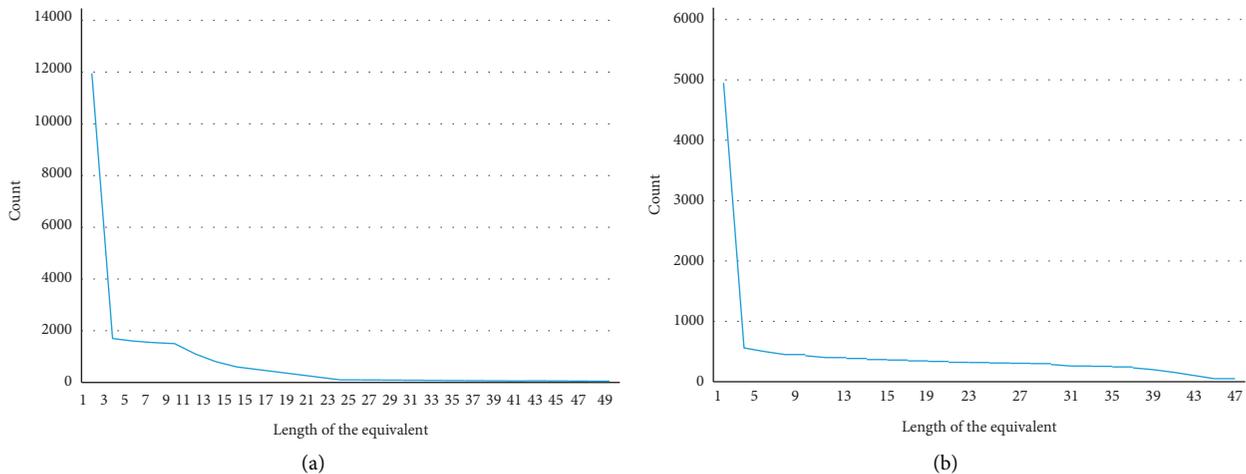


FIGURE 13: Distribution of syntactic equivalence versus node length. (a) Distribution of syntactic equivalence to node length in LDC2006T04. (b) Distribution of syntactic equivalence to node length in LDC2008E43.

language ordering errors in English-Chinese translation is longer; in other words, compared with English, there are more long-distance ordering problems in Chinese.

5.2. English Translation System Results

5.2.1. Test Results. The machine translation system runs on a PC with a P4 processor and a memory of 256 M. When the

system is running, the system occupies about 6 M memory and can process about 8000 words per minute. Judging from the translation results of the above groups of test sentences, the system can solve the ambiguity problems in sentence segmentation, word segmentation, concurrent words, ambiguous structure, and translation selection relatively well. Although this is only a targeted partial test, it cannot be concluded that the English-Chinese machine translation

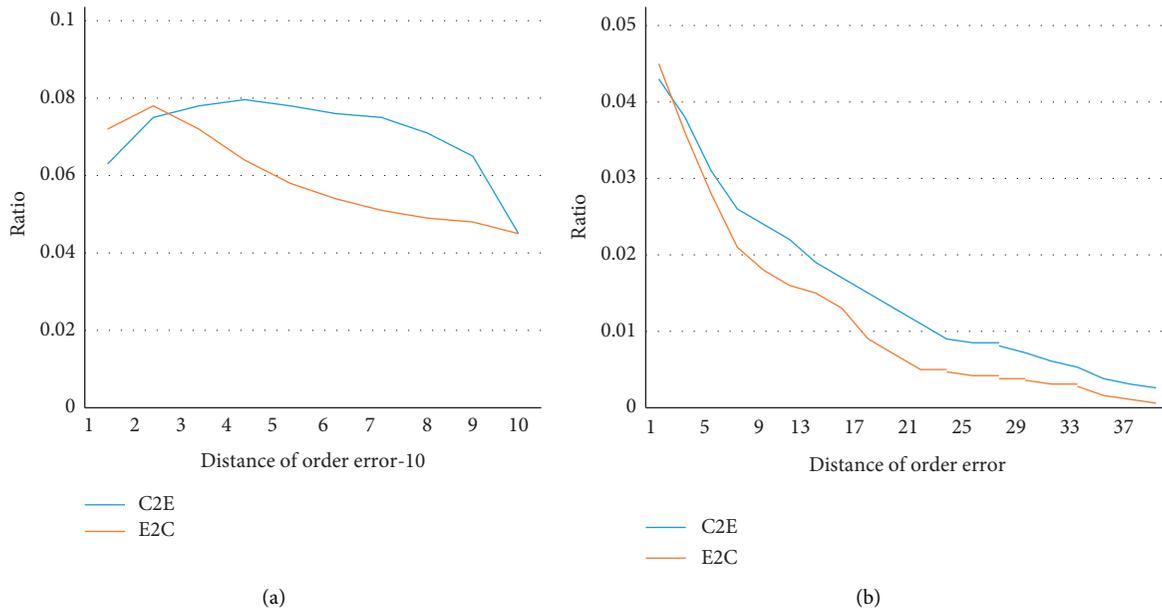


FIGURE 14: Distribution of RDt values in wrong word order transitions.

system has achieved complete success, it also partially shows the feasibility of the system's design structure and analysis algorithm.

5.2.2. Advantages and Disadvantages of the System. One of the advantages of the system design is that it divides a translation analysis process into six analysis stages before and after a translation conversion analysis unit with the number of translated clauses as a basic data. It simplifies the analysis and design steps and design implementation of the translation conversion engine function, and improves the overall translation quality accuracy and translation analysis efficiency. The coding scheme used by machine dictionaries describes the syntactic, semantic, and contextual conditions of words in an intuitive and concise manner. A rich source of knowledge provides comprehensive support for disambiguation at all stages of the translation process.

The disadvantage is that if the system is to meet the actual needs, the machine dictionary must reach at least tens of thousands of entries, and the acquisition and writing of vocabulary knowledge will become the bottleneck of the system. And the speed of operation is relatively slow, which cannot achieve the speed and precision required by the market. Because of the lack of entry, some translations will not be fully reflected in the meaning part.

6. Conclusion

The study of linguistic theory is closely linked with the study of machine translation technology. Linguistic theory is the understanding of the regularity of human natural language, and machine translation just needs this linguistic regularity. Linguistic theories at different levels, such as lexical, syntactic, semantic, and pragmatic, all play a guiding role in machine translation [29]. The current situation is that this

guiding role is not manifested significantly. For example, the midsection theory, thematic role, and phrase shift in syntactic theory are very good linguistic laws, but they have not been fully utilized in machine translation practice. Therefore, with the help of algorithms based on the Internet of Things and big data, this paper further analyzes the innovative system of English software, fills in the gaps for it, and uses another solution to correct some of the defects that cannot be expressed at the moment. In the next step, the linguistic theories most closely related to machine translation should be systematically studied and studied, and corresponding computational models should be established to guide the subsequent practice of machine translation.

Data Availability

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

Conflicts of Interest

The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

This work was supported by the Youth Project of Higher Education Research in 2021 of the 14th five-year plan of Guangdong Higher Education Society: Study on the Training Mode of Undergraduates Majoring in Translation of Application-oriented Universities in Guangdong Province under the Background of New Liberal Arts (Project No.: 21GQN88).

References

- [1] V. Tallapragada, N. A. Rao, and S. Kanapala, "EMOMETRIC: an IOT integrated big data analytic system for real time retail customer's emotion tracking and analysis," *International Journal of Computational Intelligence Research*, vol. 13, no. 5, pp. 673–695, 2017.
- [2] K. Philemon Kibiwott, Y. Zhao, Y. Zhao, J. Kogo, and F. Zhang, "Verifiable fully outsourced attribute-based sign-encryption system for IoT eHealth big data in cloud computing," *Mathematical Biosciences and Engineering*, vol. 16, no. 5, pp. 3561–3594, 2019.
- [3] H. Jamil, T. Umer, C. Ceken, and F. Al-Turjman, "Decision based model for real-time IoT analysis using big data and machine learning," *Wireless Personal Communications*, vol. 121, no. 4, pp. 2947–2959, 2021.
- [4] S. Ahmad and M. M. Afzal, "Development and effect of fog computing on the big data analysis for IoT devices in 5G era," *International Journal of Advanced Science and Technology*, vol. 29, no. 4, pp. 7033–7040, 2020.
- [5] D. J. Shin, J. M. Eun, H. G. Lee, M. G. Lee, J. M. Park, and J. J. Kim, "Big data-based log collection and analysis in IoT environments," *Journal of Engineering and Applied Sciences*, vol. 13, no. 5, pp. 1064–1072, 2018.
- [6] S. Rajeswari, K. Suthendran, and K. Rajakumar, "A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 8, pp. 365–369, 2018.
- [7] B. Zhao, X. Li, J. Li, J. Zou, and Y. Liu, "An area-context-based credibility detection for big data in IoT," *Mobile Information Systems*, vol. 2020, no. 2, 12 pages, Article ID 5068731, 2020.
- [8] J. Granat, J. M. Batalla, C. X. Mavromoustakis, and G. Mastorakis, "Big data analytics for event detection in the IoT-multicriteria approach," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4418–4430, 2020.
- [9] D. Wei, H. Ning, F. Shi et al., "Dataflow management in the Internet of things: sensing, control, and security," *Tsinghua Science and Technology*, vol. 26, no. 6, pp. 918–930, 2021.
- [10] M. Palaniswami, A. S. Rao, D. Kumar, P. Rathore, and S. Rajasegarar, "The role of visual assessment of clusters for big data analysis: from real-world Internet of things," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 6, no. 4, pp. 45–53, 2020.
- [11] K. A. Meerja, P. V. Naidu, and S. R. K. Kalva, "Price versus performance of big data analysis for cloud based Internet of things networks," *Mobile Networks and Applications*, vol. 24, no. 3, pp. 1078–1094, 2019.
- [12] A. Almeida, R. Mulero, P. Rametta, V. Urošević, M. Andrić, and L. Patrono, "A critical analysis of an IoT-aware AAL system for elderly monitoring," *Future Generation Computer Systems*, vol. 97, pp. 598–619, 2019.
- [13] K. M. Ko and S. G. Kim, "Analysis technique of a wearable IoT health information on the MAPHIS," *International Journal of Engineering & Technology*, vol. 7, no. 2, pp. 224–227, 2018.
- [14] G. Li, F. Liu, A. Sharma et al., "Research on the natural language recognition method based on cluster Analysis using neural network," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9982305, 13 pages, 2021.
- [15] M. H. o. Yi, M. J. Lim, H. Ko, and J. H. Shin, "Method of profanity detection using word embedding and LSTM," *Mobile Information System*, vol. 2021, Article ID 6654029, 9 pages, 2021.
- [16] S. Jeschke, C. Brecher, H. Song, and D. Rawat, *Industrial Internet of Things: Cybermanufacturing Systems*, Springer, Switzerland, Europe, 2017.
- [17] N. Chen, B. Rong, X. Zhang, and M. Kadoch, "Scalable and flexible massive MIMO precoding for 5G H-cran," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 46–52, 2017.
- [18] L. Chen, F. R. Yu, H. Ji, B. Rong, X. Li, and V. C. M. Leung, "Green full-duplex self-backhaul and energy harvesting small cell networks with massive MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3709–3724, 2016.
- [19] Y. Chi, D. Puthal, S. P. Mohanty, and E. Kougiyanos, "Big-Sensing-Data Curation for the cloud is coming: a promise of scalable cloud-data-center mitigation for next-generation IoT and wireless sensor networks," *IEEE Consumer Electronics Magazine*, vol. 6, no. 4, pp. 48–56, 2017.
- [20] A. K. Dogra and J. Ka Ur, "Moving towards smart transportation with machine learning and Internet of Things (IoT): a review," *Journal of Smart Environments and Green Computing*, vol. 2, no. 1, pp. 3–18, 2022.
- [21] P. Pięta and T. Szmuc, "Applications of rough sets in big data analysis: an overview," *International Journal of Applied Mathematics and Computer Science*, vol. 31, no. 4, pp. 659–683, 2021.
- [22] L. Zhang, Y. Huo, Q. Ge, Y. Ma, Q. Liu, and W. Ouyang, "A privacy protection scheme for IoT big data based on time and frequency limitation," *Wireless Communications and Mobile Computing*, vol. 2021, no. 3, 10 pages, Article ID 5545648, 2021.
- [23] F. M. Abd and M. E. Manaa, "Predicting heart diseases from large scale IoT data using a map-reduce paradigm," *Open Computer Science*, vol. 10, no. 1, pp. 422–430, 2020.
- [24] S. Ahmad, M. M. Afzal, and A. Alharabi, "Big data analytics with fog computing in integrated cloud fog and IoT architecture for smart devices," *International Journal of Information and Computer Security*, vol. 20, no. 6, pp. 171–177, 2020.
- [25] S. S. Hong, D. W. Kim, S. P. Hong, B. K. Kim, and J. K. Lee, "NoSQL-based IoT BigData fast collection and analysis system suitable for road pavement quality management," *Journal of Korean Society for Geospatial Information Science*, vol. 28, no. 4, pp. 99–107, 2020.
- [26] R. S. Apare and S. N. Gujar, "Analysis of privacy preservation techniques in IoT," *International Journal of Applied Evolutionary Computation*, vol. 10, no. 3, pp. 27–33, 2019.
- [27] P. Zhang, Y. Wu, and H. Zhu, "Open ecosystem for future industrial Internet of things (IIoT): architecture and application," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 1–11, 2020.
- [28] A. Ur, R. Butt, and S. Ahmad, "An Internet of things (IoT) based big data analytics," *Journal of Engineering Technology*, vol. 6, no. 2, pp. 70–82, 2018.
- [29] N. H. Kim, "Design and implementation of Hadoop platform for processing big data of logistics which is based on IoT," *International Journal of Services Technology and Management*, vol. 23, no. 1–2, 131 pages, 2017.