

Retraction

Retracted: English Characteristic Semantic Block Processing Based on English-Chinese Machine Translation

Advances in Multimedia

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

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

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

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

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

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

References

- [1] Y. Yu, "English Characteristic Semantic Block Processing Based on English-Chinese Machine Translation," *Advances in Multimedia*, vol. 2022, Article ID 1458394, 9 pages, 2022.

Research Article

English Characteristic Semantic Block Processing Based on English-Chinese Machine Translation

Yuxiu Yu 

School of Foreign Languages, University of Sanya, Sanya, Hainan 572000, China

Correspondence should be addressed to Yuxiu Yu; 14520184@xzyz.edu.cn

Received 27 May 2022; Revised 22 June 2022; Accepted 29 June 2022; Published 16 July 2022

Academic Editor: Qiangyi Li

Copyright © 2022 Yuxiu Yu. 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.

In order to solve the problem that the current English-Chinese machine translation software cannot understand the characteristics of English sentences repeatedly, a semantic block processing method for English-Chinese machine translation was proposed. In the process of the English sentence comprehension, English semantic block, which played an important role, was analyzed in detail. On the basis of this, the core content and characteristics of English semantic block were discussed. And with the help of the corresponding processing algorithms, taking verbs and business English as the research object, three kinds of semantic models were summarized. A lexical chunk database model based on English characteristic semantic block processing was proposed. The open test results showed that the matching success rate of the model for the semantic pattern database was about 90%.

1. Introduction

The emergence of machine translation provides powerful technical support for English-Chinese translation. The problem of machine translation research is the understanding and generation of language. And the generation of all languages lies in the understanding of language. This is also the common problem of some current machine translation software [1]. In the current stage, many machine translation software generally have the problem of sentence comprehension. Most of the software only analyzes and understands the sentence from the syntax, without in-depth consideration of the semantic connotation, so as to affect the accuracy of translation. Based on this, a language block database model is proposed in the research based on English characteristic semantic block processing to analyze the constituent characteristics of English characteristic semantic block in depth and further improve the accuracy of machine translation.

Through continuous testing of machine translation systems' processing of real text, it was found that the limited rules determined by human hands were difficult to correctly translate large-scale real sentences. The establishment of a rule-based machine translation system often required the

establishment of various knowledge bases, which needed to describe the lexical, syntactic, and semantic knowledge of the source language and the target language, and even describe the world knowledge unrelated to the language [2]. However, because knowledge bases must be created and maintained by many trained experts, it was difficult to describe and establish these knowledge bases. And with the continuous expansion of knowledge base, it became a difficult problem to ensure that the newly introduced knowledge did not contradict the old knowledge.

2. Literature Review

After the 1980s, people looked for new ways to avoid some difficulties based on rules and corpus-based methods began to appear. A corpus was a collection of real texts that were used in the real world without any modifications or embellishments to the original sentences. Corpus could be directly used for analysis, transformation, and generation in machine translation, or indirectly used as the basis for acquiring translation knowledge and statistical knowledge. The corp-based approach excluded in-depth analysis of language and attempted to conduct bilingual translation based on the mass collection of bilingual data for mutual translation.

Corpus-based methods included instance-based and statistics-based machine translation methods. Machine translation method based on the instance believed that it could find the most similar in bilingual corpus translation examples for language translation and based on statistical machine translation claimed that mathematical models were established in the translation process and used the bilingual corpus to estimate model parameters [3]. And then translation was implemented according to the model and through the estimated parameters.

For decades, researchers have explored many methods of machine translation. Some experts believed that a single method was difficult to achieve high-quality translation results. Most of the real progress in machine translation research came from hybrid methods. The combination of a corpus-based empirical approach and a rule-based rationalism approach became the consensus of many machine translation researchers in the world, and this approach had been applied to practice. For hybrid methods, there were many ways [4]. One way was the multi-engine machine translation system. Multi-engine machine translation was to integrate multiple machine translation methods in the same environment, and each machine translation engine worked simultaneously or separately. The goal of this strategy was to improve the results of the system and could be called a result-oriented strategy [5]. The system had two working modes, namely, pretranslation judgment and post-translation judgment. The working process of the pretranslation determination method was to break the same input into appropriate sentence groups, judge which translation engine each group was suitable for, and then send it to the corresponding engine, and finally combine the translations of each translation engine [6]. The working process of post-translation determination method was to send the same input to different translation engines and then select the best translation results to combine and generate the final translation after each translation engine output the translation. Another approach was the machine translation process. It was in different stages of translation using different methods to improve the accuracy of each processing stage, so as to improve the translation quality of the whole system. For example, statistical methods were used to eliminate part-of-speech and part-of-speech classes, and machine learning methods were used to learn language rules from corpus and then apply the rules to syntactic analysis. Hybrid methods combined the advantages of various machine translation methods to produce high-quality translation [7].

3. Semantic Language and Semantic Unit Theory

3.1. Semantic Language Form Definition. Sentences of different concrete natural languages can be translated into each other, and people using different concrete natural languages can communicate with each other [8], because there are sentences corresponding to the same semantics between these concrete natural languages, or a group of sentences can be established to express the semantics. A concrete natural

language consists of all semantic unit representations including all sentences, that is, all sentence-meaning representations. All semantic units constitute semantic language. Semantic language includes all sentence meaning. A concrete natural language I can be seen as a representation of a semantic language on I .

Semantic language form system is given as

$$\{A_0, \#, E_0, C, \Phi\}, \quad (1)$$

where A_0 is the terminal alphabet. $\#$ is the parameter variable identifier. E_0 is the semantic unit set (2)–(7):

$$E_0 = S_0 \cup U_0, \quad (2)$$

$$S_0 = \{a | a \in (A_0)^{**}\}, \quad (3)$$

$$U_0 = \{V(Xv) | V \in (A_0)^{**}\}, \quad (4)$$

$$Xv: ::= \#Cv^* | Xv\#Cv^*, \quad (5)$$

$$Cv^*: ::= Cv | \frac{Cv^*}{Cv}, \quad (6)$$

$$Cv \in C, Cv \subseteq C. \quad (7)$$

where C refers to the set of classes. Each semantic unit SE corresponds to a type Ce , in which

$$Ce \in C, \quad (8)$$

where Φ is the set of substitution rules for parameter variables.

3.2. Form Definition of Specific Natural Language. The form system of specific natural language is as follows:

$$\{A_1, \#, R_1, C, \Phi\}, \quad (9)$$

where A_1 is the terminal alphabet. $\#$ is the parameter variable identifier. R_1 represents the complete set for the semantic unit (10)–(13):

$$R_1 = S_1 \cup U_1 \cup V_1, \quad (10)$$

$$S_1 = \{a | a \in (A_1)^{**}\}, \quad (11)$$

$$U_1 = \left\{ \#Cv^* | Cv^*: ::= \frac{Cv^*}{Cv}; Cv \in C, Cv \subseteq C \right\}, \quad (12)$$

$$V_1 = \{\gamma | \gamma \in (A_1 \cup U_1)^{**}\}. \quad (13)$$

where C refers to the set of classes. Each semantic unit representation corresponds to one and only one type C_r , in which

$$C_r \in C. \quad (14)$$

where Φ is the set of substitution rules for parameter variables.

$\#Cv^* \rightarrow \alpha$ can be replaced under the following conditions:

$$T(\alpha) \in C_v, \quad (15)$$

$\#C_{bv}^* \rightarrow \beta$ can be replaced under the following conditions:

$$\beta \in U_1, T(\beta) \in C_v, \quad (16)$$

$\#C_{bv}^* \rightarrow \gamma$ can be replaced under the following conditions:

$$T(\gamma) \in C_v. \quad (17)$$

A sentence in a concrete natural language can be substituted by “a semantic unit representation in which all virtual quantities are replaced by a semantic unit representation without variables.” This specific definition of natural language differs from the usual definition in that there is no production and no initializer [9]. Most of the language members are objective rather than produced by the formal system, and only part of the content can be generated by variable members through variable substitution.

3.3. Machine Translation Methods Based on Semantic Units. The multilanguage machine translation system based on semantic language is composed of two parts, which is a multilanguage unified semantic unit base of high quality, complete, extensible, no discard, no repetition, no false ambiguity, and no normal ambiguity [10].

The flow chart of the multilingual machine translation system based on semantic language is shown in Figure 1.

Machine translation based on semantic language is a multinatural language machine translation method. Every natural language can be regarded as a representation of semantic language. And the translation of two natural languages can be regarded as a transformation between two representations of semantic language.

4. Analysis of Semantic Blocks of Business English Letters for Machine Translation

4.1. Theoretical Model of Chunk Formation Analysis. As a fixed structure composed of several word units, lexical chunks have entered the field of linguistics research. However, due to the differences in the research background, research purpose, observation perspective, and research methods, there are great differences in the understanding of its connotation, attribute, type, and internal systematics. At present, the scope of the research of lexical chunks is gradually expanding, and the classification of lexical chunks is becoming more and more detailed [11]. At the same time, it will inevitably bring some new problems. Because it has never been attempted to systematically investigate related phenomena and describe related concepts from the aspects of information processing units and the combination of form and meaning, it is at least chaotic. So it needs to be understood in the theoretical framework of constructional grammar. Based on the consideration of constructional grammar theory and chunk research, as well as the reference and improvement of the constructional chunk analysis [12], a theoretical model of chunk formation analysis called

“construction (chunk (word))” is proposed in the research (see Figure 2).

4.2. Definition of Business English Letter Lexical Chunks for Machine Translation. According to constructional grammar theory, chunk is a major construction. Constructions exist at different levels of language and can be regarded as constructions as long as they are paired with form, meaning, and function. Business English letter lexical chunks play the role of construction well because of their combination of ideogram and whole coding. Machine translation-oriented chunks of business English correspondence refer to those concrete communicative elements that exist as substantive “chunks” and are also the combination of form, meaning, and function.

Context can accurately activate the appropriate meaning of ambiguous words in context at any time [13]. The paradigmatic nature of lexical chunks and the matching of contextual information determine the low frequency of lexical chunk ambiguity. Lexical chunks of formulaic languages have strong contextual utility. The richness of lexical chunks information and the matching of contextual information determine that the richness of the semantic relations and semantic search space narrow. The basic context matching ensures it to be understood and extracted rapidly in a certain context, greatly improving the richness of language expression and authenticity.

The reuse degree of a language lexical chunk is the decisive factor to determine whether a language combination unit belongs to a language lexical chunk. To determine whether a language combination unit belongs to a language lexical chunk, it must be considered whether it is often used together. If the reuse degree is high, it can be identified [14].

According to the research results of psycholinguistics, cognitive linguistics, corpus linguistics and other theories and methods [15], chunk storage, and extraction as a whole can reduce the effort of language information processing. The lexical chunks of business English letters for machine translation are obtained as a whole and stored as examples in the machine translation instance database as a whole. They have relatively fixed grammatical structure and relatively stable meaning collocation, and can be extracted as a whole when needed, which not only ensures semantic integrity but also improves the speed of language output [16].

4.3. Extraction and Processing of Semantic Blocks for Business English Letters

4.3.1. Text Preprocessing Module. In the research, ABBYYAligner automatic alignment software was used to align English and Chinese business letters in parallel corpora at both discourse and sentence levels. 400 business English letters and 400 corresponding Business Chinese letters (a total of 115,026 words, including 46,868 words in English and 68,158 words in Chinese, accounting for about one-tenth of the total database) were randomly selected from the parallel corpus of English and Chinese business letters of about 1 million words [17]. Firstly, text preprocessing was

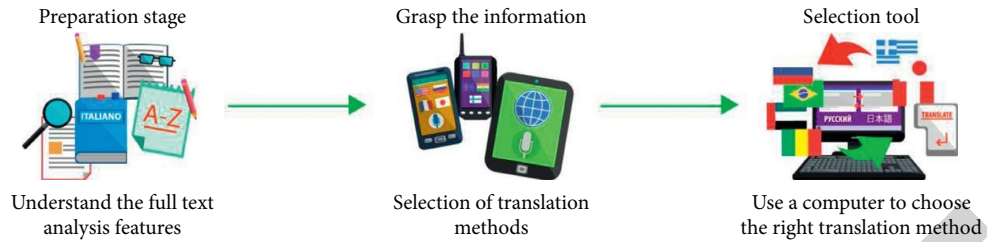


FIGURE 1: Machine translation system based on semantic language.

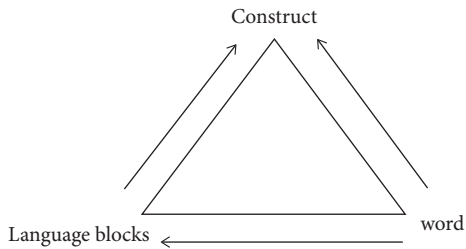


FIGURE 2: Theoretical model of lexical chunk formation analysis “construction (chunk (word)).”

carried out. 400 English letters and 400 corresponding Chinese letters were renamed. English letters were 1_E, 2_E, . . . 400_E, and Chinese letters were 1_C, 2_C. . . 400_C. Then, the format of 400 business English letters with text alignment was converted from TXT to UTF-8, because UTF-8 format was required for corpus phrase extraction system corpus input.

4.3.2. Automatic Extraction Module. In the research, 202127 Grams were automatically extracted and generated by the corpus phrase extraction system, and 154613 Grams were obtained after elimination processing. In the 150,000 Grams, Grams with a value greater than 2 were extracted according to the frequency. The MI method was used to calculate, and 290 candidate lexical chunks were obtained. The final results were imported into a word document and an Excel data table [18]. The program of the corpus phrase extraction system used big data application processing technology comprehensively, such as multi-CPU parallel computing under 64-bit platform, multithread concurrent processing and MongoDB massive data storage technology to rapidly grow n -gram sets within the specified range, and synchronous weight elimination in memory and statistics of the frequency of each Gram, left and right adjacencies, and frequencies of each adjacency [19]. The generated results were quickly stored in MongoDB through memory mapping to prepare for various operations and processing in the next step. At the same time, in the process of calculation and storage, Grams containing punctuation marks, special characters, and other obvious noise information were filtered, making the generated Grams set more pure and effective. After clean collection Grams were obtained, frequency was used for threshold selectively filter, such as 1 and 2, and then to filter results. Through the calculation method of the MI, the MI value was calculated and inclusive language chunks of

TABLE 1: Left and right adjacency structures in procedure files.

Order number	Gram	The word length of the Gram	Frequency
	List {"a": a frequency, "b": b frequency,}		
	Next word list {"a": a frequency, "b": b frequency,}		

preferential treatment were conducted. Finally, the final results of lexical chunks were obtained.

There are detailed left and right adjacencies in the procedure file, as shown in Table 1.

Based on the statistics of 290 candidate lexical chunks, the distribution of N values from 2 to 9 candidate lexical chunks is obtained (as shown in Table 2).

The following bar chart can more intuitively show the proportion of candidate lexical chunks with N values of 2–9 in the 290 automatic extraction lexical chunks (see Figure 3) [20].

According to the statistical results of the distribution of candidate lexical chunks with N values of 2–9 in the chart above, the number of candidate lexical chunks with N value 2 is 81, accounting for 27.93%, more than one-third of the total number. The total number of the four candidate lexical chunks with N value of 6–9 is 72, accounting for 24.83%, less than one-third of the total number. The total number of the four candidate lexical chunks with N values of 2–5 was 218, accounting for 75.17%, or about three quarters of the total number [21].

According to the statistical results of the frequency of 290 candidate lexical chunks, “Thank you for your letter” is the most frequent with 43 times. There are 16 cases with frequency of more than 10 times (5.5%). There are 17 cases with frequencies between 6 and 9, accounting for 5.9%, and the sum of the two accounted for 11.4%. There are 21 with 5 times frequency, accounting for 7.2%. There are 65 patients with a frequency of 4, accounting for 22.4%. The sum of the two terms is 29.6%. There are 171 cases with 3 frequencies, accounting for 59.0%, as shown in a pie chart of frequency ratios given in Figure 4.

4.3.3. Manual Processing Module. For ease of operation, the following specific criteria are listed in the research:

- (1) A solid construction, that is, a combination of form, meaning, and function
- (2) Basically unambiguous

TABLE 2: Distribution ratio of candidate lexical chunks with N values of 2-9.

N value	Quantity	Scale (%)
2	81	27.93
3	55	18.97
4	52	17.93
5	30	10.34
6	31	10.69
7	15	5.17
8	16	5.52
9	10	3.45
Total	290	100

TABLE 3: Candidate lexical chunks automatically extracted by computer ($N=2$).

N value	Quantity	Scale (%)
2	81	27.93
3	55	18.97
4	52	17.93
5	30	10.34
6	31	10.69
7	15	5.17
8	16	5.52
9	10	3.45
Total	290	100

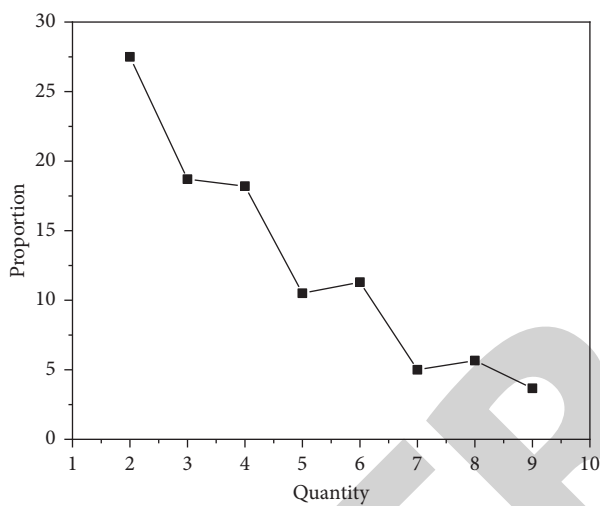


FIGURE 3: Ratio of N values of candidate chunks automatically extracted.

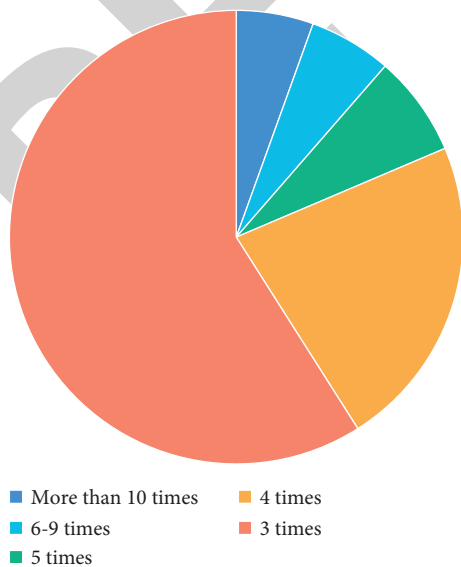


FIGURE 4: Frequency ratio of candidate chunks automatically extracted.

- (3) High degree of reuse
- (4) Lump-sum deposit and withdrawal
- (5) Multiword units

In addition to the above five items, proper nouns, trade names, and abbreviations should not be included in the machine translation-oriented business English letter lexical chunks, because they are not studied as lexical chunks in the current common definition of lexical chunks. And they are easy to find in the general dictionary and abbreviation dictionary. A specific analysis based on the N value is shown in Table 3.

Through the analysis of candidate lexical chunks $N=2$, it could be found that “New York, San Francisco, Hong Kong, paper bags, iron nails, cotton sweaters, office manager, laser printers, silk blouses, electronic products, plastic furniture, window frames, frying pan, swimming trousers” were not part of lexical chunks. CFR basis (Cost and Freight) was an acronym not included in lexical chunks [22]. The following candidate lexical chunks represented multiple meanings. These candidate lexical chunks appeared “one-to-two or one-to-many” in meaning; they were not classified as lexical chunks, such as “poor condition, sole agents, sole agency, sales promotion, debit note, claim form, less than, usual practice, good prospects, drawn up, selling season, more than, special care.” Finally, in the candidate lexical chunks with N value = 2, 14 proper nouns and commodity names, 1 abbreviation, and 13 polysemy words were excluded. The remaining 53 were included in the business English letter lexical chunks for machine translation, as shown in Table 4.

Through the analysis of candidate lexical chunks $N=3$, it could be found that “Bank of China” and “light industrial products” belonged to proper nouns and commodity names, and they were not included in language lexical chunks. “CIF New York (Cost Insurance and Freight)” and “in an FCL container (Full container Load)” were abbreviations and were not included in language lexical chunks. The following candidate lexical chunks represented polysemy and were also not listed as lexical chunks, such as “bill of exchange, bill of lading, terms and conditions, in due course, date of delivery.” Finally, in the candidate lexical chunks with N value = 3, two proper nouns and commodity names, two abbreviations, and five polysemy words were excluded [23].

TABLE 4: Candidate lexical chunks automatically extracted by computer ($N=3$).

Order number	Candidate block	Order number	Candidate block
1	Serve you well	29	Packed in cartons
2	Past four' years	30	In an fcl container
3	Make clear before hand	31	Time of shipment
4	Goods are shipped	32	Date of de lively
5	No available space	33	may rest assured
6	In fill settlement	34	In due course
7	Pleased to inform	35	Not later than
8	Date of invoice	36	Tonnes of payment
9	Prices we quoted	37	Maintain friendly relations
10	Having carefully considered	38	Very brisk demands
11	An exceptional case	39	very brisk demands

The remaining 46 were included in the business English letter lexical chunks for machine translation.

4.4. Construction Mode of Lexical Chunks Database. In the research, the construction mode of chunk database is designed. First, three layers are designed. These three layers are basic resource layer, human-computer interaction extraction layer, and corresponding rule layer (Figure 5).

The basic resource layer is mainly the parallel corpus of English and Chinese business letters, which is the basis of the construction model of lexical chunk database of English and Chinese business letters for machine translation. The human-computer interaction extraction layer mainly involves the extraction of lexical chunks of human-computer interaction business English letters. The format of business English letters with text alignment needs to be transformed to meet the needs of corpus phrase extraction system. Then, the business English letter with sentence alignment and the corresponding business Chinese letter are imported into a word document and an Excel data table, respectively, so that the business English letter lexical chunks based on the parallel corpora of English and Chinese business letters can be retrieved and the corresponding translation can be determined. The corresponding rules layer mainly uses ParaConc software retrieval query function [24]. Based on English and Chinese business letter parallel corpus for machine translation of business English letter of lexical chunks in retrieval, through the analysis for machine translation of business English letters and their corresponding Chinese translation of lexical chunks, the corresponding relationship between the corresponding rules and standards for their score is made. Finally, an English-Chinese business correspondence lexical chunk database for machine translation is formed.

The above three levels constitute the macrooperation flow of lexical chunk database construction mode. Here are six modules of the three layers, which are specific micro-operational processes. Six modules are designed in the three

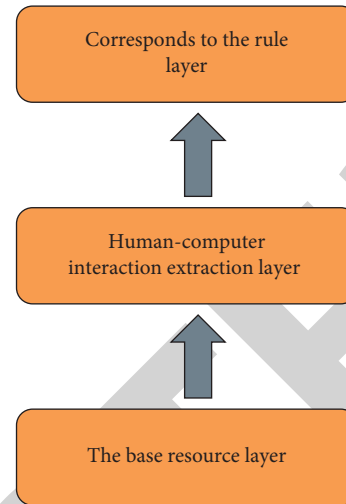


FIGURE 5: Three levels of lexical chunk database construction pattern.

levels of lexical chunk database construction mode, which are text preprocessing module, automatic extraction module, manual processing module, corresponding relation analysis module, corresponding rule formulation module, and standard score evaluation module. The human-computer interaction extraction layer includes three modules: text preprocessing module, automatic extraction module, and manual processing module. The corresponding rule layer consists of three modules: the corresponding relation analysis module, the corresponding rule formulation module, and the standard score evaluation module. In this way, the flow chart of the construction mode of the lexical chunk database of English and Chinese business correspondence for machine translation is finally formed (see Figure 6).

5. Chinese Translation Methods for Common Verbs

5.1. Selection of Chinese Translation Methods for Common Verbs Based on Semantic Patterns

5.1.1. Statistics-Based Method. The purpose of this method is to collect contextual evidence from corpus, which can be used to disambiguate ambiguous words in new input sentences. This method is mainly based on the view that the target words in the source language given the meaning of the original text are frequently selected as appropriate translation words in a similar context to the target language. A statistical method can quantitatively show the differences in linguistic phenomena, but it also needs the support of large-scale real corpus. And the data quality in the corpus will have a fundamental impact on disambiguation and translation results [25].

5.1.2. The Instance-Based Method. This method is to find the sentence most similar to the input sentence in the large-scale real corpus and make appropriate adjustments to the target

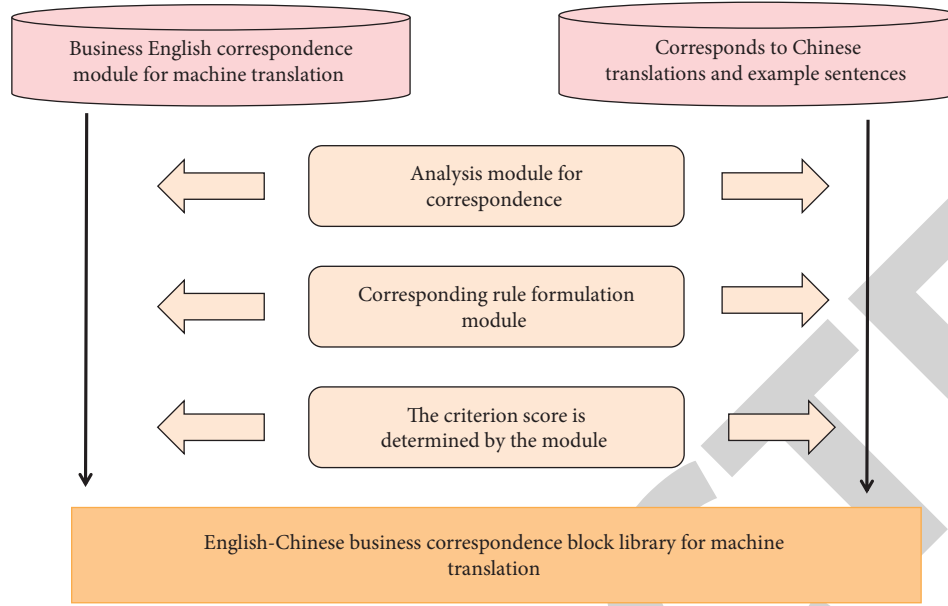


FIGURE 6: Flow chart of lexical chunk database construction mode.

language of the sentence as the output result. This method and statistics-based methods are also empirical methods. And some problems of statistical methods cannot be avoided.

Verbs are the most active part of speech in a language, and they are characterized by many changes, complex forms, and strong polysemy. Therefore, in the research, the method based on semantic patterns was adopted to translate English verbs into Chinese, so as to better adapt to the flexible polysemy characteristics of verbs.

5.2. Implementation of Chinese Translation Algorithm. In the research, semantic patterns of common English verbs were extracted. According to these semantic patterns, the semantic pattern database, fixed sentence pattern database, and variable database of 211 common verbs were constructed. A total of 8632 pairs of English-Chinese sentences and 6896 pairs of English-Chinese semantic patterns were extracted. The fixed sentence patterns related to common verbs in the text and corresponding Chinese translation were directly extracted to form a fixed sentence pattern database. There were 237 semantic types of variables related to common verbs in the variable database, as well as the specific English and Chinese expressions of each variable.

5.3. Experiment Analysis

5.3.1. Experiment Settings. Using the Chinese translation for verbs method introduced in the research, ten thousand pairs of sentences were selected from Xiamen University bilingual corpus for an open test. And the test results of randomly selected words and sentences were manually counted. Table 5 shows the basic information of the test corpus.

Table 6 shows the results of the open test.

TABLE 5: Test corpus.

The total number of sentences	Commonly used English verbs	Number of sentences containing the word
40000	Eat	99
40000	Grow	87
40000	Hear	200
40000	Kill	56
40000	Meet	161

In the table, ① indicates that the translation is correct and identical to the original Chinese translation. ② indicates that the translation result is different from the original Chinese translation, but the translation is correct after manual examination, without changing the meaning of the sentence. ③ indicates that the semantic pattern database, fixed sentence pattern database, and variable database are not matched successfully, and the translation fails. ④ indicates that the semantic pattern database, fixed sentence pattern database, and variable database are successfully matched, but the translation is wrong. The accuracy rate is used to evaluate the performance of the algorithm, which is given as

$$P = \frac{a + b}{a + b + c + d} \times 100\%. \tag{18}$$

5.3.2. Analysis of the Experiment Results. According to the experiment results, the Chinese translation method for common verbs based on semantic pattern database, fixed sentence pattern database, and variable database can achieve left and right accuracy. The reasons for translation failure are as follows:

- (1) Incomplete collection of verb semantic patterns and fixed sentence patterns in the verb semantic pattern

TABLE 6: The results of open test.

Commonly used English verbs	Occurrence number a	Occurrence number b	Occurrence number c	Occurrence number d	Accuracy (%)
Eat	83	7	7	2	90.9
Grow	42	36	8	1	89.7
Hear	156	22	17	5	89
Kill	35	16	5	0	91.1
Meet	87	53	20	1	85

database and fixed sentence pattern database leads to translation failure.

- (2) Incomplete expressions of variables in the variable database and their corresponding English and Chinese expressions lead to translation failure.
- (3) Omission and other special sentence patterns lead to translation failure.

Therefore, the accuracy of the Chinese translation of English verbs by using this method can be further improved by further expanding and perfecting the semantic pattern database, fixed sentence pattern database, and variable database. The translation errors caused by ellipsis and other special sentence patterns can also be improved by adding a context function.

6. Conclusions

The construction model of the English-Chinese business correspondence lexical chunk database for machine translation includes three layers: basic resource layer, human-computer interaction extraction layer, and corresponding rule layer. The basic resource layer is mainly the parallel corpus of English and Chinese business letters, which is the basis of the construction model of the lexical chunk database of English and Chinese business letters for machine translation. The human-computer interaction extraction layer includes text preprocessing module, automatic extraction module, and manual processing module. This layer mainly deals with monolingual language, and here is English. The corresponding rule layer consists of three modules, namely, the corresponding relation analysis module, the corresponding rule formulation module, and the standard score evaluation module. This layer mainly deals with bilingualism; here, it is from English to Chinese. In the application prospect of the research, the process of this model can be used to construct a lexical chunk database in each limited field and form a super-large lexical chunk database in all fields serving the English-Chinese machine translation system, which will definitely play a key role in improving the translation accuracy of the English-Chinese business letter machine translation system.

In the research, the translation choice of English verbs was investigated from the sentence-level bilingual corpus. Semantic patterns were used to describe the collocation rules between verbs and other words. The extraction method of semantic pattern and the construction method of semantic pattern database were introduced. Taking common verbs as the research object, the semantic pattern database, fixed

sentence pattern database, and variable database were constructed. The verb semantic pattern database was an important basis for the Chinese translation method of English verbs in the research. Currently, the verb semantic pattern database was extracted manually. Constant corrections and improvements were necessary for the amount and accuracy of extraction. In the future, it is necessary to divide verb semantic pattern more accurately and scientifically, and constantly improve the existing semantic pattern database to build a more complete verb semantic pattern database.

The experiment showed that the method based on semantic pattern database, fixed sentence pattern database, and variable database was effective. The open test results showed that the matching success rate of this model for the semantic pattern database was about 90%. However, semantic analysis and Chinese translation research are only conducted on 211 common verbs. And its scope cannot cover all verbs, so its application scope is slightly narrow. To make the method more widely used, more verbs need to be analyzed and summarized. In the future work, besides the common verbs in the research, more semantic patterns of verbs should be investigated.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the University of Sanya.

References

- [1] S. Deng and Q. Dong, "Ga-net: global attention network for point cloud semantic segmentation," *IEEE Signal Processing Letters*, vol. 28, pp. 1300–1304, 2021.
- [2] H. Cheng, J. Lu, M. Luo, W. Liu, and K. Zhang, "Ptanet: triple attention network for point cloud semantic segmentation," *Engineering Applications of Artificial Intelligence*, vol. 102, no. 5, Article ID 104239, 2021.
- [3] Q. Bi, K. Qin, H. Zhang, and G. S. Xia, "Local semantic enhanced convnet for aerial scene recognition," *IEEE Transactions on Image Processing*, vol. 30, pp. 6498–6511, 2021.
- [4] J. Ye, Y. Jing, X. Wang, K. Ou, Tao, and M. Song, "Edge-sensitive human cutout with hierarchical granularity and

- loopy matting guidance,” *IEEE Transactions on Image Processing*, vol. 29, pp. 1177–1191, 2019.
- [5] L. Zhang, J. Han, and Z. Shi, “Learning temporal relations from semantic neighbors for acoustic scene classification,” *IEEE Signal Processing Letters*, vol. 27, pp. 950–954, 2020.
- [6] R. Yin, R. Zhang, W. Zhao, and F. Jiang, “Da-net: pedestrian detection using dense connected block and attention modules,” *IEEE Access*, vol. 8, pp. 153929–153940, 2020.
- [7] F. Wang and Y. Zhang, “A de-raining semantic segmentation network for real-time foreground segmentation,” *Journal of Real-Time Image Processing*, vol. 18, no. 3, pp. 873–887, 2020.
- [8] R. Dong, X. Pan, and F. Li, “Denseu-net-based Semantic Segmentation of Small Objects in Urban Remote Sensing Images,” *IEEE Access*, vol. 7, pp. 65347–65356, 2019.
- [9] H. H. Han, L. Fan, S. Zhang, and D. Jiao, “A New Semantic Segmentation Model for Supplementing More Spatial Information,” *IEEE Access*, vol. 7, pp. 86979–86988, 2019.
- [10] X. Lan, X. Gu, and X. Gu, “Mmnet: multi-modal multi-stage network for rgb-t image semantic segmentation,” *Applied Intelligence*, vol. 52, no. 5, pp. 5817–5829, 2021.
- [11] K. Zhu, C. Xu, Y. Wei, and G. Cai, “Fast-pldn: fast power line detection network,” *Journal of Real-Time Image Processing*, vol. 19, no. 1, pp. 3–13, 2021.
- [12] P. Zhang, W. Liu, Y. Lei, H. Wang, and H. Lu, “Rapnet: residual atrous pyramid network for importance-aware street scene parsing,” *IEEE Transactions on Image Processing*, vol. 29, pp. 5010–5021, 2020.
- [13] V. K. Balasubramanian and G. R. Karpagam, “Investigation on blockchain technology for web service composition: a case study,” *International Journal of Web Services Research*, vol. 18, no. 1, pp. 1–23, 2020.
- [14] Z. Gao, L. Jiang, X. Xia, D. Lo, and J. Grundy, “Checking smart contracts with structural code embedding,” *IEEE Transactions on Software Engineering*, vol. 47, no. 12, pp. 2874–2891, 2021.
- [15] Y. Shen, H. Zhang, Y. Fan, A. P. LeeLee, and L. Xu, “Smart health of ultrasound telemedicine based on deeply represented semantic segmentation,” *IEEE Internet of Things Journal*, vol. 8, no. 23, pp. 16770–16778, 2021.
- [16] Z. Chen, Y. Fu, Y. Zhang, Y. G. Jiang, X. Xue, and L. Sigal, “Multi-level semantic feature augmentation for one-shot learning,” *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4594–4605, 2019.
- [17] W. Wu, Q. Li, Z. Chen, and S. Liu, “Semantic information oriented no-reference video quality assessment,” *IEEE Signal Processing Letters*, vol. 28, pp. 204–208, 2021.
- [18] A. M. Braoveanu and R. Andonie, “Integrating machine learning techniques in semantic fake news detection,” *Neural Processing Letters*, vol. 53, no. 2, pp. 1–18, 2020.
- [19] W. Wang, Y. Fu, Dong, and F. Li, “Semantic segmentation of remote sensing ship image via a convolutional neural networks model,” *IET Image Processing*, vol. 13, no. 6, pp. 1016–1022, 2019.
- [20] H. Yang, C. Huang, F. Wang, K. Song, and Z. Yin, “Robust semantic template matching using a superpixel region binary descriptor,” *IEEE Transactions on Image Processing*, vol. 28, pp. 3061–3074, 2019.
- [21] A. Sharma, R. Kumar, Talib, S. Srivastava, and R. Iqbal, “Network modelling and computation of quickest path for service-level agreements using bi-objective optimization,” *International Journal of Distributed Sensor Networks*, vol. 15, no. 10, Article ID 155014771988111, 2019.
- [22] J. Jayakumar, B. NagarajNagaraj, S. ChackoChacko, and P. Ajay, “Conceptual implementation of artificial intelligent based E-mobility controller in smart city environment,” *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1–8, Article ID 5325116, 2021.
- [23] J. Liu, X. Liu, J. Chen, X. Li, T. Ma, and F. Zhong, “Investigation of ZrMnFe/sepiolite catalysts on toluene degradation in a one-stage plasma-catalysis system,” *Catalysts*, vol. 11, no. 7, p. 828, 2021.
- [24] R. Huang, S. Zhang, W. Zhang, and X. Yang, “Progress of zinc oxide-based nanocomposites in the textile industry,” *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 44, pp. 281–289, 2021.
- [25] Y. Zhang, X. Kou, Z. Song, Y. Fan, M. Usman, and V. Jagota, “Research on logistics management layout optimization and real-time application based on nonlinear programming,” *Nonlinear Engineering*, vol. 10, no. 1, pp. 526–534, 2021.