

## Research Article

# Design and Testing of Automatic Machine Translation System Based on Chinese-English Phrase Translation

Jing Ning  and Haidong Ban

*College of Foreign Language, Xijing University, Xi'an 710123, Shaanxi, China*

Correspondence should be addressed to Jing Ning; 20120080@xijing.edu.cn

Received 29 June 2021; Revised 2 September 2021; Accepted 9 September 2021; Published 30 September 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Jing Ning and Haidong Ban. 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 development of linguistics and the improvement of computer performance, the effect of machine translation is getting better and better, and it is widely used. The automatic expression translation method based on the Chinese-English machine takes short sentences as the basic translation unit and makes full use of the order of short sentences. Compared with word-based statistical machine translation methods, the effect is greatly improved. The performance of machine translation is constantly improving. This article aims to study the design of phrase-based automatic machine translation systems by introducing machine translation methods and Chinese-English phrase translation, explore the design and testing of machine automatic translation systems based on the combination of Chinese-English phrase translation, and explain the role of machine automatic translation in promoting the development of translation. In this article, through the combination of machine translation experiments and machine automatic translation system design methods, the design and testing of machine automatic translation systems based on Chinese-English phrase translation combinations are studied to cultivate people's understanding of language, knowledge, and intelligence and then help solve other problems. Language processing issues promote the development of corpus linguistics. The experimental results in this article show that when the Chinese-English phrase translation probability table is changed from 82% to 51%, the BLEU translation evaluation system for the combination of Chinese-English phrases is improved. Automatic machine translation saves time and energy of translation work, which shows that machine translation shows its advantages due to its short development cycle and easy processing of large-scale corpora.

## 1. Introduction

People express their emotions through language, which is an important tool for communication between people. Therefore, it is more and more important to overcome communication barriers between languages in the 21st century. Machine automatic translation is a meaningful and complicated research and full of challenges and difficulties. Continuous and high-quality automatic translation machine research is one of the ultimate goals of computing and language research, which is the main trend of future development.

Machine automatic translation is becoming more and more important in today's society, and its potential is huge with the rapid economic development. Everyday people from all walks of life deal with a large number of documents,

and people of different languages communicate with each other. Therefore, machine automatic translation has a great market demand, and only a very large amount of information can meet the needs of translation. With the combination of Chinese and English, automatic machine translation has become the most common method at present, which has the benefit of greatly facilitating people's lives, and it is also a simple data warehouse.

Sangeetha and Jothilakshmi proposed a speech-to-speech translation system, which mainly focuses on translation from English to Dravidian. The three main technologies involved in the SST system are automatic continuous speech recognition, machine translation, and text-to-speech synthesis systems. Based on automatic associative neural network, vector support mechanism, and hidden Markov model, automatic continuous speech recognition has been

developed. Compared with SVM and AANN, HMM produces better results, but it currently lacks specific data to prove [1]. Shereen and Mohamed believes that deaf-mute people are an important part of the growing community, and they use sign language. However, communication between normal people and hearing-impaired people becomes difficult because most normal people cannot understand the meaning of sign language gestures, while deaf-mute people cannot understand natural spoken language. There are approximately 70 million deaf and hearing-impaired people in the world, as well as people who use sign language as their mother tongue or mother tongue. The analysis of the existing system provides us with the necessary information about its work process, success rate, shortcomings, and limitations, and its development is relatively vague [2]. In order to improve the accuracy of automatic machine translation, Sangeetha and Jothilakshmi proposed a study to improve the efficiency of machine translation when necessary. For this reason, based on the adjustment of English context and the mutual information between words in English words, they proposed an automatic translation system based on semantic relations [1].

The innovation of this article lies in the investigation and study of the method probability of Chinese-English phrase translation and the combined machine automatic translation system of Chinese-English phrase translation. The systematic research and experimentation of automatic translators are of great significance. To a certain extent, it can promote the rapid and in-depth dissemination of international information.

## 2. Machine Translation Method of Chinese-English Phrases

**2.1. Process of Machine Translation of Phrases.** Machine translation experiments found that the translation model in the basic IBM machine translation equation was replaced by the reverse translation model, but the accuracy of translation was not reduced by the automatic translation machine, which could not be passed through the channel theory [3, 4]. Therefore, the maximum entropy based on machine translation is proposed. This more general method is a statistical method of machine translation based on the source channel [5, 6]. Characterizing the maximum entropy, language format and translation mode, and adding them to the model framework, the main advantage is that it can easily integrate knowledge sources and automatically weight between knowledge sources. Most current statistical machine translation methods use the highest entropy modeling framework [7, 8]. The automatic machine translation modeling of phrases is shown in Figure 1.

### 2.2. Method and Process of Machine Translation Based on Phrase Structure

**2.2.1. Corpus Preprocessing.** The processing level of the corpus directly affects the translation results. Statistical machine translation usually uses a bilingual corpus and prepares Chinese and English corpora separately [9, 10]. The results of corpus preprocessing are shown in Table 1.

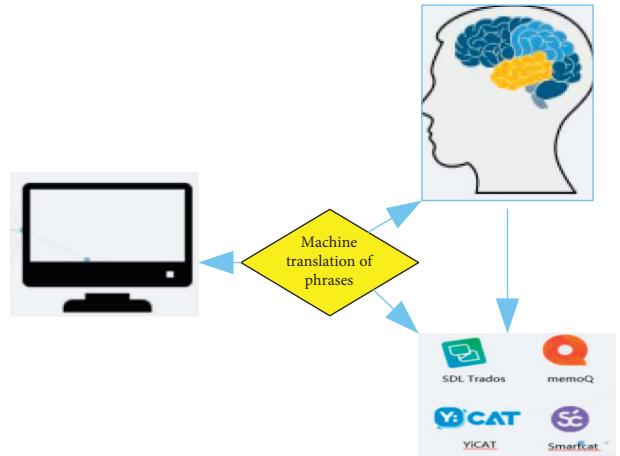


FIGURE 1: Automatic machine translation modeling of phrases.

TABLE 1: Experimental data.

Step	Complaint text
1	Introduce custom dictionary
2	Segment words and mark part of speech
3	Remove stop words
4	Keep important part of speech words
5	Keep processing results

**2.2.2. Implementing the Title Translation System in the Aviation Field.** On the basis of researching related machine translation theory, using some existing resources and tools, we complete the phrase translation model module, realize the phrase-based statistical machine translation system, and introduce the basic working principle of the system, system implementation, and system operating environment settings and parameters [11, 12].

**2.2.3. Automatic Evaluation Technology of Machine Translation.** Based on the research of machine automatic translation technology, the results of Chinese-English machine automatic translation are automatically evaluated. In the field of statistical machine learning, there are already some methods to solve domain adaptation problems [13, 14]. But most of them are only used to solve simple learning problems (such as classification or regression). In the face of structured learning problems such as machine translation, different domain adaptive methods are used to solve them separately under the machine learning framework [15]. The application of machine translation automatic evaluation technology is shown in Figure 2.

**2.3. Stack Search Translation Method.** Stack search utilizes a research and exploratory method. Before strengthening the search of  $n$  heaps, the number  $n$  is the number of words in the source language sentence, and each state data hypothesis is stored in the stack extension [16]. “I” is translated as “she,” and “flower” is derived from the word “flowers.” Both hypotheses are in the first stack of the stack search translation method [7]. Also, in the second stack, the molecule adds

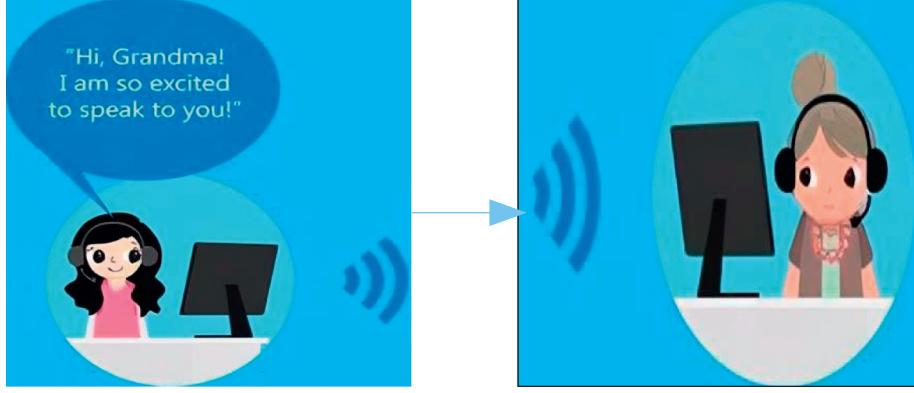


FIGURE 2: Application of machine translation automatic evaluation technology.

information about the source language translation of the two terms. For the source language words that have been translated, the stack cost is low, so they are determined as the best translation [8]. The stack search conversion is shown in Figure 3.

### 3. Chinese-English Machine Translation Probability Experiment

**3.1. Phrase-Based Statistical Machine Translation.** The basic idea is to use phrases as the basic unit of translation. In the process of transfer, everyone's translation of phrases is not the same everywhere, and there are various opinions and interpretations at the same time. In grammatical sense, if only the phrase lines are not continuous, we still need to solve the problem of the overall coherence of the full text. In order to expand the transmission of these contents, we can easily solve the local problem in the same way. Context-dependent issues and explanations of phrases in all languages using this method can maintain the original state of the language to the greatest extent. Generally speaking, the so-called free grammar method can be a continuous line subnavigation. Therefore, Chinese-English translation of words must be carried out to extract the viewpoint of double-body protection, and the process of rule-based machine translation is shown in Table 2.

**3.2. Defining the Format of Phrase Translation Probability Table.** In the output file of the phrase output module, each line contains some Chinese phrases, English phrases, and translation probability values:

$$P(\vec{b}, \vec{a}) = \frac{N(\vec{b}, \vec{a})}{\sum_t N(\vec{b}, \vec{a})}. \quad (1)$$

Lexicalized translation probability:

$$\text{lex}(a^j, b^j, c) = \prod_{jm}^j \frac{1}{\left| \left\{ i(j, i) \in c \mid \sum_j p(a_j, b_j) \right\} \right|}. \quad (2)$$

The BLEU evaluation tool is currently the most widely used indicator in international machine translation

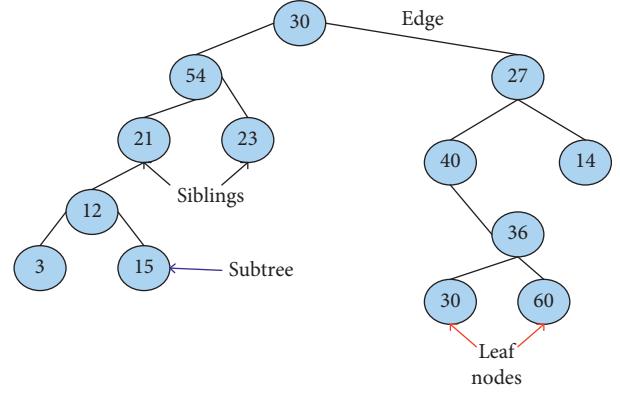


FIGURE 3: Stack search conversion.

TABLE 2: Rule-based machine translation process.

Step	Translation	Knowledge base
1	Source language text	Target
2	Lexical and morphological analysis	Rule base
3	Syntax analysis	Dictionary
4	Semantic analysis	Target language text
5	Structure transformation	Language generation

evaluation. It compares the system translation with the reference translation, calculates the accuracy of each system translation, and finally records the entire translation. It is calculated as follows:

$$\text{sore} = \text{BP} * \exp \left( \sum_{n=1}^N f_n \log h_n \right) W_i. \quad (3)$$

**3.3. Vector Machine Algorithm.** Where  $P$  is the penalty length factor,  $B$  is the shortest length of the reference translation of the tested sentence, and  $R$  is the translation length of the tested sentence, that is, the number of words contained in the entire output translation.

$$BP = \min \left\{ 1, \exp \left( 1 - \frac{R_{ref}}{R_{sys}} \right) \right\}. \quad (4)$$

In the current statistical method, the shared modernity of indecent words indicates the fidelity of translation. It means that a word has been translated in the original text, and a dictionary with more than two yuan appears at the same time to indicate the fluency of the target language:

$$H_n = \frac{\sum_{u=1}^u \delta_n}{u = 1 \sum_u a_n}. \quad (5)$$

The way and form of this formula equal to half is calculated as follows:

$$\text{score} = \frac{2 * p * r}{p + r}. \quad (6)$$

This is the minimum error rate during editing. The score ranges from 0 to 1. The scores are different for editing. The so-called edit distance is the minimum cost of insertion, deletion, and replacement operations performed by converting the system output into a reference translation:

$$P = \frac{M}{L}. \quad (7)$$

**3.4. Automatic Evaluation Model of Machine Translation.** The logarithmic linear model is introduced into statistical translation, which can add any number of features to the translation process and determine the contribution of each feature to the translation result by weighting these features. Therefore, the effect of phrase-based translation system developed by them is far better than that of word-based translation system. For formal syntax model rules, the formula is as follows:

$$\begin{aligned} F &= F\left(\frac{a}{y}\right) = \frac{\text{count}(a, y)}{\sum_a \text{count}(a^i, y)}, \\ F &= F\left(\frac{a}{r}, a\right) = \prod_{i=1}^n \frac{1}{\{j / (i, j) \in a\}}, \\ F &= F\left(\frac{r}{a}\right) = \frac{\text{count}(a, y)}{\sum_r \text{count}(a, y^i)}. \end{aligned} \quad (8)$$

According to CKY algorithm, we can construct hypergraph from sentences of source language. When we calculate the  $k$ -best derivation of a node, the ranking of the dimension of rules no longer only depends on the score of syntactic rule features. We use heuristic function  $H(R)$  to sort the rules:

$$\begin{aligned} h(Y) &= u \sum_{i=0}^r InP(w_i) - u(r+1), \\ f(a \cdots a_l) &= \prod f_{lm} \left( \frac{a_i}{a \cdots a_{i-1}} \right). \end{aligned} \quad (9)$$

## 4. Chinese-English Phrase Translation Combined Machine Automatic Translation System

**4.1. Phrase-Based Statistical Machine Translation.** The basic idea is to use machine translation as the basic unit of phrase translation. In the translation process, each translated word must be combined with context and constrained translation during the translation process. But generally speaking, no grammar is performed in the same way. In this way, the two-body alignment should be removed from the bilingual excerpt. Given a source language sentence, the sentences used for the translation process model are as follows: the source is divided into phrase sentences and language word viewpoints, and the order is adjusted according to the interpretation target model of each sentence. Phrases are used as the basic unit of translation. The Chinese sentence interpretation system is used to divide many sentences into so-called “phrases” and then translate them into English. The generated phrases and output are shown in Table 3.

**4.2. Translation Process.** It mainly includes the following parts: model phrase translation, translation model training, language training, and trial transmission of decoding results. These parts are scattered in the form of a flowchart. From the perspective of the translation science model, each table is best to learn Chinese phrases from the English interpretation of English sentences and arrange them in a row as shown in the flowchart in Figure 4.

Traditional word alignment-based heuristic phrase extraction methods will have word alignment errors and word-to-space problems, which leads to the loss of many bisyntactic phrases. On the other hand, the bilingual phrases extracted from bilingual phrases in this paper are bilingual phrases with better quality. Therefore, we consider adding the extracted bisyntactic phrases to the phrase table to make up for the bisyntactic phrases lost by the heuristic phrase extraction method. The experiment uses the provided training set, development set, and test set. The source language of the corpus is Chinese, and the target language is English. The scale of the experimental data is shown in Table 4.

It can be seen from the table that English sentences are on average longer than Chinese sentences, and both Chinese and English sentences are longer, especially when the average length of English sentences reaches one word, which brings difficulty to syntactic analysis. We analyze the syntax of the source language and the target language and extract bilingual phrases using an iterative phrase extraction algorithm. According to the Chinese-English phrase translation training set, there are 120,000 Chinese-English bilingually aligned sentences, and the test corpus contains 141 sentences. In the experiment, this paper uses the  $C$  value and the degree of adhesion to reduce the source language. It is added to the translation model as a function, and the translation results are compared with the reference frame. First, no matter how long the sentence is, the possibility of translation is lower than the  $C$  value of the source code, and it can be

TABLE 3: Example of phrase-based statistical machine translation process.

Original	She will attend the party on May 1st		
Phrase division	She will	On May 1st	Attend the party
Translation	She will	On May 1st	Attend the party
Adjusting the order	She will	Attend the party	On May 1st

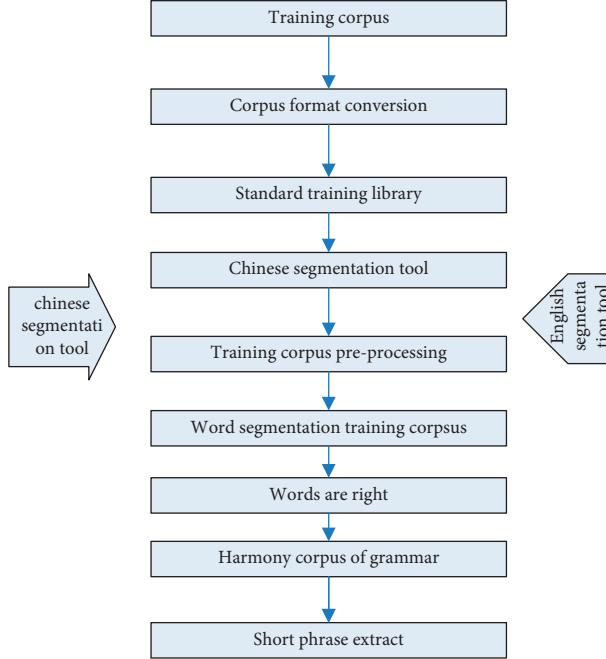


FIGURE 4: Phrase translation model training process.

TABLE 4: Experimental data.

	Chinese	English
Sentences in the training collection	99933	99777
Number of words in training set	245699	348766
The training set's average sentence length	25	30
Number of sets	400	400
Development set's number of single words	1012	1122
Test set's average sentence length	25	33

seen that the BLEU evaluation can be improved by 0.02 at most compared with the benchmark system, while the phrase translation probability table is only 78% of the original. When the phrase translation probability table is reduced to 51% of the original, the BLEU evaluation is still slightly higher than the benchmark system. The experimental results are shown in Figure 5.

The input of the bilingual phrase extraction algorithm is an aligned Chinese-English bilingual tree, so it is necessary to perform syntactic analysis on the source language end and the target language end of the training corpus separately. The bilingual phrase extraction algorithm extracts bilingual phrases based on word alignment. The training corpus is the training corpus that has been word aligned, so it is no longer necessary to apply word alignment to the training corpus. We run the bisyntactic phrase extraction algorithm and

temporarily store the extracted bisyntactic phrases. This experiment needs to run four different machine translation systems. These systems are statistical machines based on bisyntactic phrases that are generated after the extracted bisyntactic phrases are applied to the system.

**4.3. Chinese-English Translation Corpus.** The Chinese-English translation corpus is used. This corpus contains more than 10,000 words and 10,000 pairs of sentences. This article finds that the best translation effect can be achieved by using it as a means of word alignment extraction in Chinese-English translation. Used as the evaluation standard, the calculation script adopts the standard script. This article uses ten thousand sentence pairs in ten thousand pairs of sentences as the training corpus, and the number of short sentence

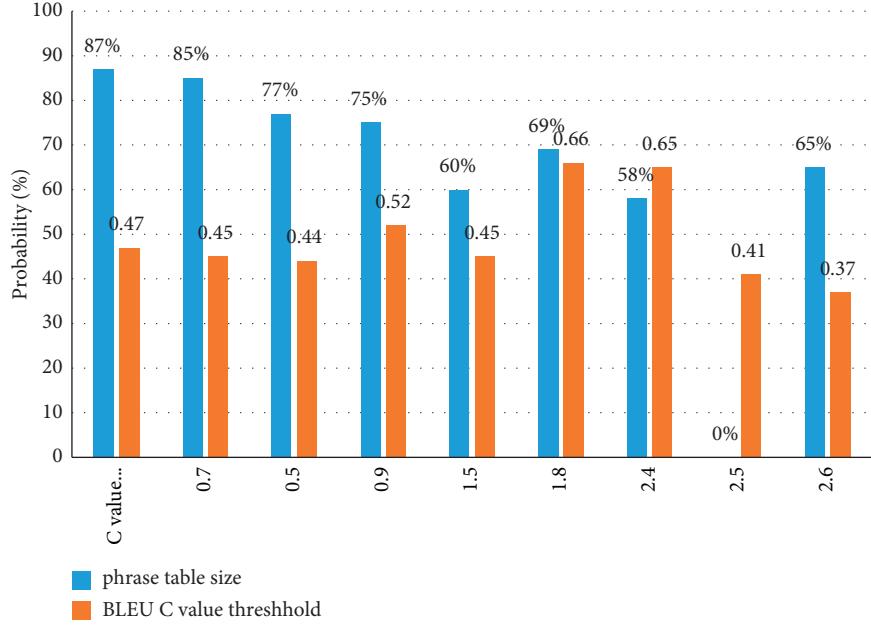


FIGURE 5: Phrase translation probability table size and BLEU value under different C value thresholds.

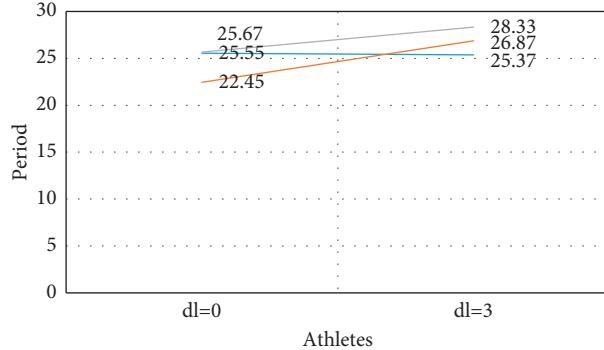


FIGURE 6: The result of the improved system under Chinese-English translation.

TABLE 5: Dataset characteristics.

Dataset	Chinese		English	
	Sentence number	Word number	Sentence number	Word number
Training set	3000	230	4000	4455
Test set	4000	456	7000	788
Development set	490	2788	3891	342

pairs extracted by the method is regarded as the parameter in the linear rearrangement model, as shown in Figure 6.

As shown in Figure 6, the fragment probability phrases are used in the Chinese-English translation, thereby improving performance. Through the data test of an example, after completing the translation process, machine translation is introduced into the system, and a partition system is established. The model and module are given, and the existing local resources and document resources are used

TABLE 6: The characteristics of phrase model and formal syntax model.

Model	Phrase model
Maximum rule length	20
Model number	334K
Model size	12K
Before filtering	34 MB
After filtering	76 MB

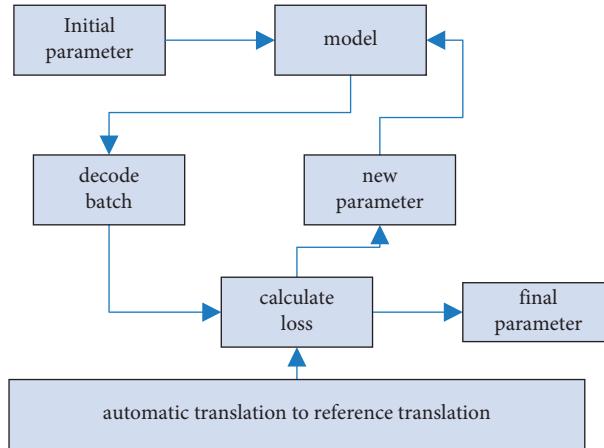


FIGURE 7: Training of machine translation.

TABLE 7: The paradigm of machine translation.

Intermediate language	Transformation method	Direct translation method
1	Source language deep representation	Target language deep representation
2	Source language text	Target language text
3	Chinese	English

including some open source translation tools and publicly authorized translation tools. These tools are based on the research and development of the comprehensive decision-making mechanism of the statistical system.

**4.4. Model Training and Parameter Setting.** The evaluation of machine translation mainly includes manual evaluation and automatic evaluation. The advantage of manual evaluation is high accuracy, but the disadvantage is that the labor cost and time cost are too high. The advantages of automatic evaluation are low cost, fast speed, and the ability to be used repeatedly. The disadvantage is low accuracy. At present, the focus of machine translation evaluation research is how to improve the rate of automatic evaluation. The test set of CSTAR 2003 is the development set of the experiment. Some features of the corpus are shown in Table 5.

The phrase is extracted from the training set, and the English part of the training set is trained by language model tool. The feature model of phrase model and formal syntax model is reduced by a 3-element language model; in order to speed up the training of minimum error rate and save memory space, the development set and test set are used to filter these models. The characteristics of the model are shown in Table 6.

The evaluation of machine translation plays an important role in the research of machine translation technology and the promotion of market. Manual evaluation refers to the evaluation of candidate translations given by machine translation system according to certain standards and norms. Automatic evaluation is the use of machines to complete the scoring process, but it requires that the results of scoring are consistent as much as possible with the person's score; the training of machine translation is shown in Figure 7.

FIGURE 8: Neural machine translation system (<http://alturl.com/acrv0>).

Machine translation evaluation, in short, is the evaluation of all aspects of machine translation in order to correctly and objectively reflect the achievements and functions of machine translation. The significance of machine translation evaluation is to find out the problems existing in the research and development of machine translation system by evaluating the performance and development level of machine translation, define the goal, find solutions, provide direction for the improvement of the existing machine translation system, and constantly improve the translation quality of machine translation system; the paradigm of machine translation is shown in Table 7.

Machine translation is a reliable way to evaluate the performance of a translation system. However, it usually takes time and effort to organize a manual evaluation. The use of automatic evaluation tools can greatly reduce the cost of evaluation, analyze the system performance in time,

improve the system targeted, and shorten the product development cycle; the neural machine translation system is shown in Figure 8.

## 5. Conclusions

This article extends the discussion from the perspective of automatic translation systems for mechanical design. The machine translation system is a large-scale system composed of several modules, which can complete the translation work. This article makes full use of the existing resources and tools in the literature, briefly describes the phrase and probability of phrase translation, and integrates these tools and modules, and we believe that building a machine translation system based on statistical results means an attempt that cannot be done by learning translators. Automatic machine translation is a complete process that integrates the development of concepts, opens up the use of existing resources, and adds modules such as repositories, dictionaries, and so on. The decision is based on the results of statistical machine translation methods that can achieve better translation results.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was supported by the Scientific Research Program funded by Shaanxi Provincial Education Department (grant no. 18JK1188) and the Scientific Research Foundation of Xijing University (grant nos. XJ180113 and XJ130134).

## References

- [1] J. Sangeetha and S. Jothilakshmi, "Speech translation system for English to dravidian languages," *Applied Intelligence*, vol. 46, no. 3, pp. 534–550, 2017.
- [2] A. Shereen and A. Mohamed, "A cascaded speech to Arabic sign language machine translator using adaptation," *International Journal of Computer Application*, vol. 133, no. 5, pp. 5–9, 2016.
- [3] J. Zhang, Y. Zhou, and C. Zong, "Abstractive cross-language summarization via translation model enhanced predicate argument structure fusing," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 10, pp. 1842–1853, 2016.
- [4] Z. Qin, P. Wang, J. Sun, J. Lu, and H. Qiao, "Precise robotic assembly for large-scale objects based on automatic guidance and alignment," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 6, pp. 1398–1411, 2016.
- [5] H. A. Bouarara, R. M. Hamou, and A. Rahmani, "BHA2: bio-inspired algorithm and automatic summarisation for detecting different types of plagiarism," *International Journal of Swarm Intelligence Research*, vol. 8, no. 1, pp. 30–53, 2017.
- [6] D. Tolic and S. Hirche, "Stabilizing transmission intervals for nonlinear delayed networked control systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 488–494, 2017.
- [7] D. Kuehn, M. Schilling, T. Stark, M. Zenzes, and F. Kirchner, "System design and testing of the hominid robot charlie," *Journal of Field Robotics*, vol. 34, no. 4, pp. 666–703, 2017.
- [8] S. F. Rafique, J. Zhang, M. Hanan, W. Aslam, A. U. Rehman, and Z. W. Khan, "Energy management system design and testing for smart buildings under uncertain generation (wind/photovoltaic) and demand," *Journal of Tsinghua University: English Edition*, vol. 23, no. 3, pp. 254–265, 2018.
- [9] C. Wang, "Design and research of ultrasonic nondestructive testing system for conveyor belt," *Machinery Management Development*, vol. 33, no. 1, pp. 98–100, 2018.
- [10] J. Li, S. Yang, H. Zhang, G. Liu, and T. Sun, "Design and field testing of a nitrogen circulation drilling system," *Chemistry and Technology of Fuels and Oils*, vol. 53, no. 3, pp. 428–435, 2017.
- [11] H. Totoki, Y. Ochi, M. Sato, and K. Muraoka, "Design and testing of a low-order flight control system for quad-tilt-wing UAV," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 10, pp. 2423–2431, 2016.
- [12] L. Yang and W. Li, "Design and implementation of indoor environment testing system based on android platform," *Environmental Science and Management*, vol. 42, no. 5, pp. 26–29, 2017.
- [13] S. Lei and Z. Liping, "Design and implementation of automatic testing system for LTE-M based TAU," *Electronics World*, no. 14, pp. 40–41, 2017.
- [14] S. Zhou, D. Zou, and T. Xiao, "Design and experiment of the velocity-pressure characteristic testing system for seafloor sediments," *Ocean Technology*, vol. 36, no. 5, pp. 55–61, 2017.
- [15] Y. Cheng, X. Chen, and H. Wang, "Design and precision analysis for PLC-based energy efficiency testing system of electric fans," *Journal of Testing Technology*, vol. 30, no. 1, pp. 1–5, 2016.
- [16] Y. Xu, W. Haikun, and S. Fang, "The design of the testing system of the diesel generator under the low temperature and low pressure," *Electrical Automation*, vol. 38, no. 3, pp. 85–87, 2016.