Optimization of English Automated Translation Depending on Weak Grammar Rules

XianJun Liu*
Guangdong University of Science & Technology, China, 523083

*Corresponding author e-mail: liuxianjun@163.com

Abstract. Machine translation was first produced in the 1940s, it with birth of computer technology and a new subject, the purpose of machine translation is to use computers as a platform for natural language translation, but for complex languages, machine translation is still lacking. Most of the language mechanisms formed by machine translation are crude and rely too much on the manual summary of language experts. In this way, it is easy for translation errors to occur in machine processing languages. In this case, how to effectively improve English machine translation under weak grammar rules has become a major problem that has not been solved in the computer field. This paper firstly analyses the principle of English machine translation and the improvement method of English machine translation under the weakening of grammar rules, and proposes an improvement and optimization method of English machine translation under the weakening of grammar rules.

Keywords: Weak Grammar Rules, English Machine Translation, Automatic Optimization

1. Introduction
Machine translation refers to the use of computers to automatically convert one human language into another human language. It is an important branch of natural language processing, and the realization of this great goal has been the dream of mankind for a long time since the birth of computers [1-2]. In the early stage of machine translation, rule-based methods were mainly adopted which emphasized the grammatical regularity of natural language and formed the processing mechanism of natural language. However, for the complex and changeable natural language, this method has many disadvantages. First, the acquisition of rules is usually manually summarized by language experts, which requires a lot of manpower[3-4]. Secondly, the granularity of rule knowledge described by language experts is relatively coarse, which is not conducive to the processing of some details of the language. Finally, the system based on rule processing has a high translation error rate. In recent years, some scholars have
done more research on data-driven methods. They include the statistical method and the instance-based method, both of which use corpus as the source of translation knowledge and do not need a complete description of natural language grammar, and both of them have good adaptability. However, such methods need a large corpus instance library as a support, and the construction of a comprehensive instance library costs a lot, and in the case of a large instance library, the problem of efficient query also needs to be solved. Both methods have their own advantages\[^{[5-6]}\]. The advantage of rule-based translation method is that useful information can be obtained by analyzing and reasoning grammar rules. The advantage of the instance-based translation method is that it can get accurate information directly from the instance library. Therefore, the combination of data-driven empirical methods and rule-based theoretical methods can often achieve better translation results.

2. Principle of English machine translation
In the process of the improvement of machine translation of English, to lose one source statements of semantic analysis, using the given in both Chinese and English grammar transformation rules, access to the source statements phrase and the corresponding English phrase links, on the basis of the collection of English phrases to make use of semantic block translation selection combination, and the output of the target language sentences, complete machine translation in English. The specific steps are as follows:

Assuming that \( T \) represents the source statement set, \( T_i \) represents the source statement after segmentation, and \( j \) represents the translatable type of the source phrase sequence, formula (1) is used to obtain the source statement phrase and the corresponding English phrase link:

\[
D_y = \frac{R}{D_n} \cdot (T \times T_i)_j,
\]

where \( R \). Represents the given Chinese-english grammar conversion rule, and \( D_n \) represents the semantic block.

Assuming that \( \hat{\alpha}(t) \) represents the main semantic role and \( p(o) \) represents the recognition of prepositional phrases, Eq. (2) is used to select and combine the English phrase set in translation:

\[
S_{w_e} = \frac{w_j \cdot \hat{\alpha}(t) \cdot SC(o)}{p(o)}
\]

In the formula, \( W_j \) represents the modifier of the word, and \( SC(o) \) represents the number of semantic blocks constructed according to the grammar library.

Suppose that \( \Phi^* \) represents the extracted feature of the source sentence, \( H \) represents any entry in the dictionary, and \( \gamma^* \) represents the value of the entry in different grammatical formats, then use Eq. (3) to improve the English machine translation:

\[
WE_{\gamma} = \frac{\Phi^* \cdot H(F) \cdot S_{w_e}}{D_{\gamma^*}}
\]

In the above formula, \( u \) represents the redundancy of knowledge.
The above methods can show that the improvement principle of English machine translation can effectively complete English machine translation. However, most English machine translation relies too much on the manual summary of language experts to obtain the grammar rules, so it has the disadvantage of large errors in semantic feature extraction.

3. System design
The basic architecture of the translation prototype system with weak grammar rules is shown in the figure 1. The system mainly consists of three modules: preprocessing, translation engine and output results.

![System Architecture Diagram](image_url)

**Figure 1.** The basic architecture of the translation prototype system with weak grammar rules

1) Pre-processing module: divide the text to be translated into sentences; eliminate abbreviations from sentences; Remove too many Spaces in a sentence.
2) Translation engine module: this module is the core part of the system. It is mainly composed of four parts: syntax analysis, semantic block recognition, grammar rule library and semantic block library. When a sentence is received by the translation engine module, it is delivered to the syntax analysis section. The syntax analysis section USES the grammar rule base to analyze the features of the sentence, mark the various parts of the sentence, and then it is processed by the semantic block analysis section, which USES the semantic block library to identify the semantic blocks in each part of the sentence.
3) Translation output module: use the processing results (syntactic rule information and semantic block information) of the translation engine module to translate and output the translation results. Below we give the specific processing process:
   1) Input the text to be translated.
   2) Divide the input text into several sentences and place them in an array. Then take a sentence and translate it.
   3) First, preprocess the sentence (such as eliminating abbreviations and removing too many Spaces). Scan the sentences to be translated from the beginning, identify the characteristics of the sentences, match the corresponding syntactic rules of the sentences in the rule base, mark the components of the sentences, and decompose the sentences into their components. If no matching
syntactic rules are found, the whole sentence is treated as a part.

4) Semantic block matching is carried out for each component part with entity meaning in the sentence, and the implicit part of speech between semantic blocks is used to deduce the larger semantic units, up to the whole of the sentence elements, are synthesized into a semantic unit.

5) After the synthesis of all the components of the sentence is completed, the translation output of the translation engine starts to work. It adjusts the order between the components of the translated target sentence according to the bilingual grammar rules, and modifies the translation to make it more consistent with the habits of the target language.

6) Take out the next sentence and go to 3).

7) End and output.

4. Library construction
Since both the syntactic and semantic block analysis sections interact with the syntax rule library, and the semantic block analysis sections also interact with the semantic block library, we must first build a semantic block library and a grammar rule library. Just as a person's experience plays a role in his own thinking, these two libraries play a similar role in our system. Here's a brief description of how to build these two libraries.

4.1. Construction of semantic block library
The method of instance-based translation requires a huge collated translation instance (bilingual text) as the source of translation knowledge. We also need a lot of organized bilingual text in our system, but these texts are mainly semantic chunks (words or phrases with actual and independent meanings) and do not require strict bilingual alignment. Currently, there are many dictionaries available, and there is an enormous amount of resources on the Internet, where we can get bilingual texts from websites that support bilingualism. However, the directly obtained bilingual text needs some processing before it can be put into the semantic block library. We need to plan the storage of semantic blocks. In multiple semantic block libraries, one of the main semantic block libraries stores some semantic blocks that people often use in daily life, and the other semantic block libraries are semantic block libraries of various specialized fields. Such division is beneficial to the disambiguation of translation and the accuracy of translation.

4.2. Construction of the grammar rule library
Rules of grammar if master language experts concluded that search engines use these rules to tag sentences, thus the recognition of semantic block has a lot of help, moreover translation engines use grammatical rules to recognize speech quickly in the library to synthesis of semantic chunk, the final translation engines use grammar rules to implement, implement never translation text into the target language conversion.

5. Simulation proof
An experiment is needed to prove the effectiveness of the proposed optimization method for English machine translation under weak grammar rules. An optimized platform for English machine translation under weak grammar rules is built on Windows platform. The experiment extracts 800
grammatically standard Chinese sentences from the official website of People's Daily as test data for translation. First, the professional English teacher gives a complete and standardized translation of the experimental Chinese test sentences, and then evaluates the translation of the machine translation system optimized by different methods based on the translation given by the teacher.

5.1. Comparison of accuracy of English machine translation with different algorithms
The algorithm in this paper and the original algorithm are used to optimize the English machine translation. The word error rate and text segmentation error rate of the two algorithms in English machine translation were compared.

The comparison results are shown in Figure 2 and 3.

![Figure 2](image_url)

**Figure 2.** The error rate of words in translation was compared by different algorithms

![Figure 3](image_url)

**Figure 3.** Segmentation error rate of translation by different algorithms
Analysis can be obtained from the Figure 2 and 3, the algorithm presented in this paper to optimize English machine translation accuracy is higher than the original algorithm, this is mainly because English machine translation in the algorithm presented in this paper, optimization of process, the use of a good system design, effectively guarantee the accuracy of this algorithm to optimize English machine translation.

6. Conclusion

English machine translation is a processing mechanism of natural language, but most of the English machine translation is too dependent on the manual summary of language experts to obtain the grammar rules, which leads to large errors in the process of English machine understanding under some weakened rules, as well as big errors in the extraction of semantic features. Therefore, an optimization method of English machine translation with weak grammar rules is proposed.

References

[1] Rootte. Research on semi-automatic construction method of Tibetan semantic dictionary based on WordNet [J]. Journal of Xizang University (natural science edition), 2014, 29(2): 48-53.
[2] Zhao danpei, et al. Identification of airport and oil depot targets based on significant semantic model [J]. Journal of Computer-aided Design and Graphics, 2014, 26(1): 47-55.
[3] Zhang Ming, Wang yan, Lu qing. A fast face retrieval method based on semantic features [J]. Journal of Chongqing Technology and Business University (natural science edition), 2014, 31(2): 56-60.
[4] Liu qi, Xiao yanghua, Wang wei. A typical classification relation identification method for mass Chinese text [J]. Computer Engineering, 2015, 41(2): 26-30.
[5] Zhu Mingwu, et al. Object recognition method based on local contour feature in natural scene [J]. Computer Engineering and Application, 2016, 52(1): 162-167.
[6] Hoover C, Sommer H. Automated Translation of Chinese-to-English Creative Literature[J]. 2010.