English Machine Translation System Based on Semantic Selection and Information Features

Pengyan Lu

Nanchang Vocational University, Nanchang City, Jiangxi Province, 330500, China
**Corresponding author's e-mail: 469421416@qq.com*

Abstract

The development of machine translation system has not yet fully realized automatic and high-quality, because human language is complex, and there are certain limitations in understanding language rules. Therefore, in the research of machine translation, it not only is limited to the syntactic and semantic analysis of a single sentence, but also requires an in-depth analysis of the internal laws of the language, including: sentence groups, paragraphs, chapters, and contextual information inherent in genres. The English machine translation system based on semantic selection and information features conducts in-depth analysis of the source language sentences according to the semantic unit library, and completes the translation in combination with the target language sentences. A multi-language machine translation system based on semantic language needs a unified multi-natural language machine translation software and a multi-language unified semantic unit library (the basis for the establishment of the machine translation system).

Keywords: *computer technology; semantic selection; artificial intelligence; information analysis; English translation*

1 INTRODUCTION

On the basis of analyzing and arranging sentences, they are classified according to the relationship between English tenses and Chinese time expressions [1]. The sentences include: Chinese-English bi-sentences and Chinese sentence times, etc [4]. This method can reduce the complexity and intersection of tense classification. In the process of proposing these concepts, it is necessary to formalize the information, establish a time information base, and propose many Chinese sentences, which are combined by a variety of strategies [5]. They include: Chinese single sentence time analysis algorithm, Chinese relative word marker sentence time analysis algorithm, quasi-subjunctive mood sentence time analysis algorithm, etc. It also provides a theoretical basis for the study of machine translation systems based on semantic languages [6].

2 OVERVIEW OF MACHINE TRANSLATION RESEARCH

Machine translation has a long history [7]. The French scientist Alcuni GB proposed the idea of machine translation as early as the 1930s [12]. In the 1940s, when computers appeared, humans began to use computers to carry out the idea of natural language translation. British engineer ADBooth and American engineer W. Weaver put forward the idea of machine translation with the help of computer in 1949 through long-term research and exploration. According to the rules method, language knowledge such as rules description, internal structure rules of natural language, etc. is used, and computer translation is guided by rules. It can be mainly divided into direct translation method, intermediate language method and conversion method [2].

Direct translation method, also known as word-by-word translation method. The so-called direct translation method starts from the surface sentences of the source language and converts words or fixed phrases into the target language. The conversion process is the direct translation method [3]. This method allows the computer to translate word by word without doing too much analysis of the source language. Therefore, the method 's understanding of the translation process is still not perfect.

Transformation method: Translation is carried out by means of internal expressions, and the three stages are carried out in sequence. First, the source language is converted into an internal representation of the source language. Next, the internal representation of the source language is converted into the internal representation of

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 963-967, 2023.
https://doi.org/10.2991/978-94-6463-040-4_145

the target language. Finally, the internal expression of the target language is generated into the target language. And, these conversions are performed in bilingual. Many systems currently use this structure [8].

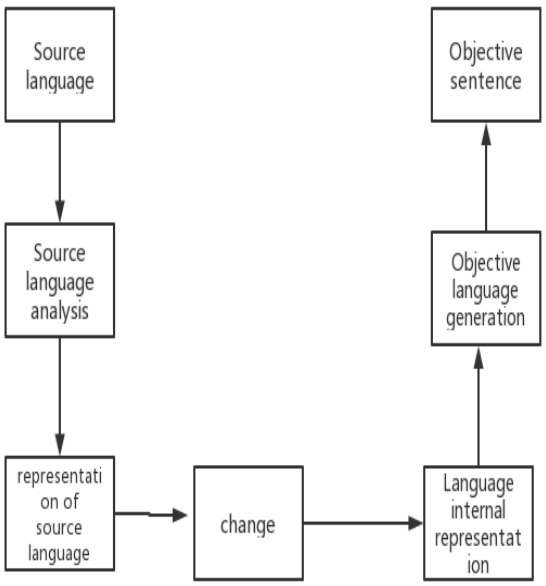

Figure 1: Transformation-based machine translation system

3 RESEARCH ON SEMANTIC DISAMBIGUATION AND CHINESE TRANSLATION OF ENGLISH PREPOSITIONS

The basic part of the machine translation system based on semantic language is the semantic unit library, which has the characteristics of high quality [9], expandability, no repetition, and no abnormal ambiguity. However, up to now, the semantic unit library applied to the system is still not perfect, and the current semantic unit only concentrates on the extraction of multi-sentence sub-level corpus in general and special contexts. English prepositions have the characteristics of particularity and flexibility [10]. Therefore, it is necessary to extract semantic units for these phrases, so as to improve the existing semantic unit library. Because the original semantic unit library is insufficient in the extraction of preposition semantic units, the semantic analysis and translation errors of prepositions increase, so whether the translation of prepositions is correct or not has a great impact on sentences [14].

4 MULTILINGUAL MACHINE TRANSLATION

A multi-language machine translation system based on semantic language needs a unified multi-natural language machine translation software and a multi-language unified semantic unit library (the basis for the establishment of the machine translation system). The semantic unit library has the following characteristics: scalable, high-quality, complete, without false ambiguity, without discardable or abnormal ambiguity [13].

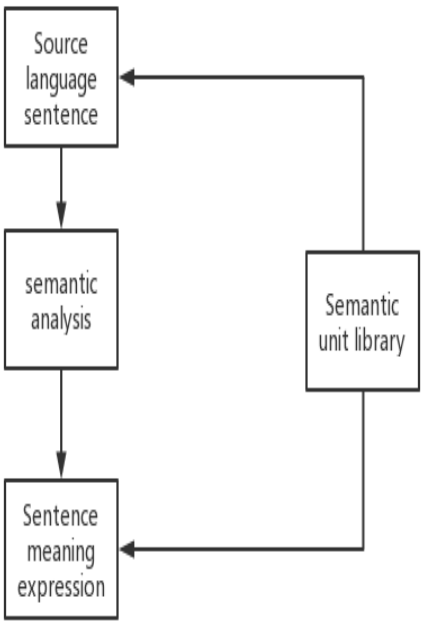

Figure 2: Machine translation system based on semantic language

For example, the translation process of the English sentence "Mr. Smith puts water into his cup" is as follows:

The process of semantic analysis is: Mr.Smith puts water into his cup→Mr.Smith puts Water into his cup→M (Smith) puts Water intoHis (Cup)→Put...into (M (Smith), (Water) His(Cu))1 to generate the semantic expression "Put...nt (M(Sith), (Water), His(Cup))", or marked as "" where 1, 2 and other numbers are the number (ID) of the semantic unit.

The process of Chinese sentences is: 1, 45(6) →1(2(Smith), water, 5(teacup) →1(Mr. Smith, water, his teacup) →Smith The gentleman poured water into his teacup [9].

It is clear that the features of semantic unit-based machine translation are as follows:

The translation of N natural languages does not belong to the NN-1 set and the 2N set based on the intermediate language . It only needs to establish N sets of translation systems and semantic analysis research.

Since no intermediate language is required, the translation process takes only one translation time, and is

not based on two translation times of the intermediate language. The operation of the process is simple, and it belongs to a replacement process according to the replacement rule.

5 SYSTEM APPLICATIONS

The application process of the formal quantifier selection algorithm in the machine translation system is as follows:

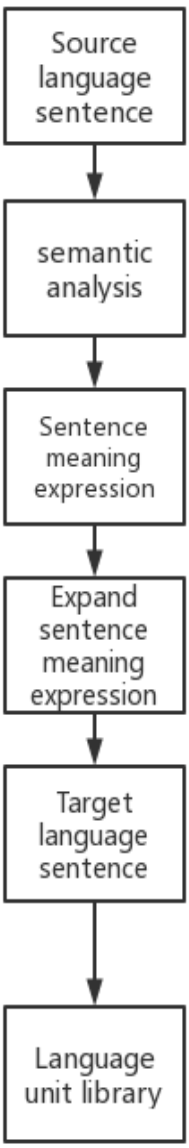

Figure 3: System application flow chart

6 EXPERIMENTAL COMPARISON

In order to verify the effectiveness of the English translation system proposed in this study, this study made an experimental test comparison, comparing the translation results of the machine translation system based on semantic selection and information features and the traditional machine translation system based on phrases [11].

In the machine translation based on semantic selection and information features, the experiment selects an English corpus, which has a large number of English materials, and conducts English translation based on these. The system will identify English phrases and grammars in the English corpus to construct a training corpus.

The results of the comparative experiments are shown in Table 1.

Table 1: Comparison of the accuracy of machine translation results based on semantic selection and information features and that of traditional phrase-based machine translation systems

system	Accuracy
Phrase-based traditional translation systems	69.3%
	71.4%
	63.8%
Machine Translation System Based on Semantic Selection and Information Features	89.3%
	96.5%
	91.1%

It is obvious from the table that the translation results based on semantic selection and information feature translation system are more accurate. The system proposed in this study has stronger syntactic analysis ability. This experiment can prove the validity of this study.

7 CONCLUSION

The research of this paper focuses on the English-Chinese sentence-level bilingual corpus, and analyzes the English machine translation system based on semantic selection and information features . Some systems are combined, so that the translation quality of the system is improved to a certain extent, and a preposition-specific semantic unit preposition semantic pattern library is further constructed. Therefore, there is a close relationship between the system's translation quality and the preposition-specific semantic unit preposition semantic library, and they influence each other. The translation process of prepositions is to perform semantic analysis on the English phrases and sentences containing prepositions in the semantic pattern database, so as to obtain the complete English semantic pattern, and then expand it through analysis, substitution, etc., and finally

obtain the Chinese complete semantic pattern and Chinese translation. According to the semantic analysis stage, three preposition-specific semantic analysis methods are obtained, which are based on link grammar and semantic pattern method, semantic pattern decomposition method and semantic pattern expansion method.

REFERENCES

[1] AllenJ. Natural language understanding [M]. Translated by Liu Qun, Zhang Huaping, Luo Weihua, etc. Beijing Electronic Industry Press 2005.

[2] Brown PF Cocke J PietraS D, et al. A statistical approach to MT [J]. Computational Linguistics 1990, 16(2):79-85

[3] Yu Shiwen. Modern Chinese Grammar Information Dictionary [M] Beijing: Institute of Computational Linguistics, Peking University, 2000.

[4] Feng Zhiwei. Research on Machine Translation [M]. Beijing: China International Translation and Publishing Company, 2004.

[5] Gao Qingshi, Gao Xiaoyu, Hu Yue Semantic-based machine translation system and method [P] China, invention patent 0113168962001.

[6] Gao QS, Hu YLi Let al.Semantic language and multi-language MT approach based on sL[J]. Journal of Computer Science & Technology, 2003, 18(6):848-852.

[7] Hu Yue, Gao Xiaoyu, Gao Qingshi. Formation method of high-quality semantic unit library in multilingual machine translation system [J] Journal of University of Science and Technology Beijing 2008, 30(6): 698-703

[8] HaroldL. Somers, Current research in MT [J]. Machine Translation 19937(4): 231-246.

[9] Hutchins J. Machine translation: past, present, future [M]. England: Ellis Horwoods Limited, 1986.

[10] JurafskyD, MartinJH. A Survey of Natural Language Processing [M] Feng Zhiwei, Sun Le, etc. Translated by Beijing Electronic Industry Press 2005.

[11] NagaoM.A framework ofa mechanical translation between Japanese and English by analogy. principle [M]. In: Elithorn A, BanerjiReds. Artificial and Human Intelligence. NATO Publication 1984:173-180.

[12] Wang Xiaojie, Chang Baobao. The foundation of natural language processing technology [M] Beijing: Beijing University of Posts and Telecommunications Press, 2002.

[13] Yao Tianshun, Zhu Jingbo, Zhang Li. Natural Language Understanding: A Study of Making Machines Understand Human Language [M] (2nd Edition) Beijing: Tsinghua University Press, 2002.

[14] Zhao Tiejun. Principles of Machine Translation [M]. Harbin: Harbin Institute of Technology Press, 2001.