

# 中国的翻译技术：过去，现在和将来

冯志伟

语言文字应用研究所 北京朝内南小街 51 号, 100010

电话: 0086-10-6525-0063

电子邮件: [zwfengde@public.bta.net.cn](mailto:zwfengde@public.bta.net.cn)

**内容提要:** 本文回顾了中国的翻译技术的发展, 对于将来机器翻译的发展提出了一些建议: 与术语数据库技术相结合以实现专业术语处理、与语料库加工技术相结合以提高原语分析的水平、与语音技术相结合以实现口语自动翻译、与文字识别技术相结合以实现文本自动输入、与译文存储技术相结合以避免重复的翻译、与网络技术相结合以克服 Internet 上语言障碍。

**关键词:** 机器翻译, 翻译技术, 译文存储

## Review of Translation Technology in China: Past, Present and Future

Feng Zhiwei

State Language Commission of China, Chaonei Nanxiaojie 51, 100010 Beijing, China

Tel: +86-10-6525-0063,

Email: [zwfengde@public.bta.net.cn](mailto:zwfengde@public.bta.net.cn)

**Abstract:** In this paper, the author gives a brief review for the translation technology in China. Some suggestions for future development are proposed. They are as follows: combination with terminology data bank in order to process technical terms, combination with corpus linguistics in order to improve the analysis level of source language. combination with speech technique in order to realize speech automatic translation. combination with OCR technique in order to automatically input the text of source language. combination with TM (Translation Memory) technique in order to reuse the translated results, combination with Internet technique in order to overcome the language barrier.

**Key Words:** machine translation, translation technology, translation memory.

### I. Past

The study of machine translation (MT) in China started over forty years ago. The past development of MT in China can be described in four periods:

- the early experimental period (1956-1966)
- the stagnant period (1966-1975)
- the recovery period (1975-1987)
- the blossom period (since 1987)

In 1956, the MT research has been included in the National Plan for Developing the Science and Technology as a project named "machine translation -- establishing

of the translation rules of natural language and mathematical theory for natural languages". This project can be divided to two parts: one is the establishing the translation rules of natural language -- "machine translation", another is the studying of mathematical theory for natural language -- "mathematical linguistics" (the theoretical foundation for machine translation). Several research groups were founded in Beijing (Academia Sinica, Beijing Institute of Foreign Language), in Guangzhou (South China Polytechnic Institute) and in Harbin (Harbin Polytechnic University). At 1959, a Russian-Chinese MT experiment (RC-59) has been successfully realized on a general-purpose computer 104. With a vocabulary of 2,030 Russian words, an algorithm of 29 flowcharts, this RC-59 experiment encouraged the belief that MT from foreign language into Chinese is possible. In the same time, some Chinese scientists began to study the MT from English to Chinese. At 1960, an English-Chinese MT algorithm was composed. A specialized brochure "Preliminary of Machine Translation" was published in 1965.

However, from 1966 until 1975 the MT in China was stagnated completely.

Since 1975, after a long sleeping of 10 years, the MT in China restarted and came to the recovery period.

At November of 1975, a MT joint-research group was established. This MT group consists of the Institute of Scientific and Technical Information of China (ISTIC), the Linguistics Institute of Chinese Academy for Social Sciences (CASS), the Computational Technique Institute of Academia Sinica, ...etc. This group carried out a MT experiment from English to Chinese on the TK-70 computer and T-4100 information processing device of Chinese characters. The raw materials contain 9,200 English titles of scientific and technical papers. As the results of this MT experiment, a MT system TITLE-1 was set up in 1986.

In same time, MT study was carried out in many institutions. For example, in Helongjiang University (Harbin), in the Mars Institute (Beijing), in the Telecommunication University (Beijing), in South China Polytechnic University (Guangzhou), in Central China Polytechnic University (Wuhan), in the Institute of Scientific and Technical Information of Shanghai (Shanghai), and in Inner Mongolia University (Huhehot).

In this recovery period, interactive approaches and diverse strategies were developed, some multilingual systems have appeared, and the application of AI technique in MT began to be considered. The mathematical linguistics was also studied in the universities or the institute. A monograph "Mathematical Linguistics" was published in 1985.

For the sake of investigation of linguistic phenomena, a English corpus was created in Jiaotong University (Shanghai), and a lot of Chinese corpora were created in Wuhan University, Peking University, Qinhua University, and Shanxi University.

In recovery period, most of MT system is experimental:

- TITLE-1 system: English-Chinese, ISTIC, 1976-1986.
- ECMT-1 system: English-Chinese, Linguistics Institute, CASS, 1978.
- JFY system: English-Chinese, Linguistics Institute, CASS, 1976-1984.
- FAJRA system: Chinese-French/English/Japanese/Russian/German, Institute of Scientific & Technical Information of China (ISTIC), 1981.
- INSPEC system: English-Chinese, Telecommunication University, 1985.
- HT-83 system: English-Chinese, Helongjiang University, 1983.
- RI-84 system: English-Chinese, Helongjiang University, 1984.
- GCAT system: German-Chinese, Applied Linguistics Institute, SLC (State Language Commission), 1985.
- FCAT system: French-Chinese, Applied Linguistics Institute, SLC, 1985.

The TITLE-1 system possessed a large-scale electronic dictionary including a basic dictionary (20,000 entries) and an idiomatic dictionary (67,000 entries). This system can translate the English titles of scientific papers in the field of metallurgy to Chinese, the average translation speed is 80 titles/hour in TK-70 mini computer (64K memory). In that time, the speed is not slow.

Since 1986 the MT of China came to the blossom period. The symbol of this blossom period is the KEYI-1 English-Chinese system of Mars Institute (Beijing), At March of 1987, KEYI-1 system has passed the academic appraisal by experts. The experts said, the translation ability of KEYI-1 can compare with the ability of the graduated students of English department in China, its translation speed is 3,000 words/hour, and the result of translation is readable. In the process of machine translation, the user can input their special words to KEYI-1 system in order to adapt to their special demands.

KEYI-1 system quickly became an operational system and was commercialized, China National Software & Technology Service Co. (CS&S) bought the copyright of the system and KEYI-1 system was renamed as TRANS-STAR system. The company CS&S puts it on the market and gains the profit.

Now TRANS-STAR system was improved and it became much better than KEYI-1. The translation speed is raised to 15,000 words/hour for 286 PC, 30,000 words/hour for 386 PC. The basis dictionary includes 40,000 entries, the system has 10 specialized technical dictionaries including 350,000 entries. The subject fields involved computer, economics, telecommunication, ceramics, thermal power industry, printing machine industry, automobile & tractor industry, Petroleum prospecting, geology, Chemical industry.

The research work of all these MT systems was supported by Chinese government funding. The goal of the machine translation is just for translation of scientific documents in order to exchange the scientific information with developed countries

(Firstly with Soviet Union, then with western countries and USA). It is why Chinese government was interested for translation technology. By this reason, there is not any private company that has interest for MT system in this period.

## II. Present

Recently, following TRANS-STAR system, many commercial MT systems are put forward to the market.

-- GAOLI MT system (English-Chinese): This MT system was developed jointly by Beijing GAOLI Computer Co. Lid. & Linguistics Institute of CASS.

- . Basic lexical dictionary: 60,000 entries in which usage and grammatical function of every word is described in detail.

- . Linguistic rules: more than 800 rules used for syntactic analysis of English and generation of Chinese.

- . Background knowledge database: more than 150 entries used for semantic analysis and generation.

- . Translation accuracy: 80%

- . Readability of translated text: 80%-90%

-- 863-IMT/EC system (English-Chinese): This MT system was developed by the Institute of Computer Technology, Academia Sinica. Now it was commercialized and got very good economic benefits.

- . Basic English lexical base: 35,000 entries

- . Basic Chinese lexical base: 25,000 entries

- . Linguistic rules: 1500 rules

- . Translation accuracy: 80%

-- SINO-TRANS system (Chinese-English): This MT system was developed by the Company CS&S (China National Software & Technology Service Co.) at 1993.

- . Basic dictionary: 40,000 entries

- . Two special subject technical dictionaries: Navel ships and boats (9312 entries), rocket-gun (33,773 entries)

- . Linguistic rules: 1,000 rules

--TONGYI system (English-Chinese): This MT system is developed by the Tianjin DATONG computer software company.

- . WINDOWS platform

- . Different special subject dictionaries:

- a. commonly-used scientific terms: 200,000 entries

- b. the terms including 22 different subjects (ex. machine building, telecommunication, aviation, medicine, etc): 3000,000 entries

- . Good market strategy and service

- . Cooperation with enterprises

- YIWANG system: This MT system was developed by SUNSHINE company of Shenzhen.
  - . Highest translation speed: 100 sentences per second.
  - . Can be used for browse the text of INTERNET.
  - . Multi-windows display.
- YIBA system: This system was developed by YAXINCHENG software technical company.
  - . Three translation models: on line translation model, automatic translation model and interface translation model
    - . Open to users: user can revise the dictionary and rules in MT system
    - . Rich special subject dictionaries: 30 subjects (ex. Computer, telecommunication, medicine)
- E-to-J system (English-Japanese): This MT system was developed by JEC company in Beijing.
  - . Technique of transformation from phrase tree (P-tree) to dependency tree (D-tree).
  - . Closely integrated with word processor

The main methods of machine translation of China are as follows:

- MMT model (Multi-branch Multi-label Tree model)
- IC analysis (Intermediate Constituent analysis)
- LS method (Logic-Semantic method)
- ST method (String Transformation method)
- I-Tree method (Integrated Tree method)

Since 1989, the corpus approach (ex. statistical approach, example-based approach) is introduced to machine translation, all the research works of machine translation are based on the processing of large-scale authentic corpus. The combination of machine translation with corpus approach will promote the development of the language translation technology in China. The corpus linguistics plays more and more role in the language translation technology in China. Some MT systems are based on the example in the corpus. Following are some Example-Based MT systems in China (EBMT):

--Japanese-Chinese EBMT system: This EBMT system is developed in computer department of Qinghua university at 1996.

-- DAYA EBMT system: This EBMT system is developed in Harbin Polytechnic University. Now EBMT is just a test. There is not the commercial EBMT system in China.

In this period, the support of Chinese government is increasing. The funds come from the different sides:

- . The national fund for social sciences: In linguistic section, MT research was supported since 1997.

- . The national fund for natural sciences: In information science section, MT research was supported always.

- . The Hi-Tech 863 fund: 863-IMT/EC system (English-Chinese), SUNSHINE YIWANG

system was supported by 863 fund.

. The 905 Chinese Language Processing Project: This project was completed in 1998.

Many private companies are interested with the research & development of MT systems. For example, GAOLI MT system is supported by GAOLI private computer company, YAXIN MT system is supported by YAXINCHENG private software company, TONGYI MT system is supported by private software computer company.

Who is the user of MT product? I believe there are three types of user:

- Government: in order to exchange with other countries.
- Common people: MT software steadily becomes the popular software that is necessary for common people, because the demand for translation grows with each passing day. The MT market was formed, and the private companies play more and more important role in driving the MT market.

Most of the user is the person with middle level of English language.

- The state large-scale and medium-scale enterprises: they are potential users of MT product. In these enterprises, there are many technical documents need to be translated into Chinese, but the document is huge, in this case, MT rough translation texts can be welcome by these enterprises.

### III. Future

New strategies will be introduced in translation technology:

- MT software combines with terminology data bank.
- MT software combines with technique of language corpus processing.
- MT software combines with speech technology in order to develop speech-to-speech translation system.
- MT software combines with Chinese characters recognition technology in order to input the source language text automatically.
- MT software combines with the Technique of Translation Memory (TM) in order to develop the practical translation technology.
- MT software combines with INTERNET technique.

Now a new project (973) start. The 973 project is directly supported by Chinese government. It will support the creation research in Natural Language processing including machine translation. Now an automatic speech-to-speech translation system (English-Chinese) is developing in Institute of Automation of Academia Sinica under the support of 973 project.

The future of the language translation technology will be more and more brilliant.

#### Reference

1. Feng Zhiwei, automatic translation, Knowledge Press, Shanghai, 1987.
2. Feng Zhiwei, Machine Translation of natural languages, Language Publishing House, 1995.
3. Feng Zhiwei, Natural Language Processing by Computer, Foreign Language Teaching Press, 1996.