

Machine Translation in Korea

Se-Young Park

Electronics and Telecommunications
Research Institute
Korea

Gil-Rok Oh

Electronics and Telecommunications
Research Institute
Korea

Abstract

This report introduces the current situation of machine translation in Korea. Recently, the need for further developing machine translation has been generally recognized. Although a few machine translation softwares for Korean have been released on the market, they do not sufficiently meet the need of users. As a result, the Korean machine translation field is only a niche market. However, several projects are underway in Korea which include world-wide technical cooperation. This report surveys the history of machine translation in Korea and describes the current market, R&D status, and current technical difficulties.

1 Introduction

There are common two responses made when we suggest a project concerning machine translation to information and telecommunication R&D sponsors. First they ask, "Are you still working on a project for machine translation?". Second, they state "It cannot be done." There responses are quite natural, because machine translation projects have been carried out for many years, but no quality products have been marketed.

However, there have been many erroneous ideas popularized about machine translation. There has been no continued and full R&D in machine translation in Korea over the last 10 years. R&D spending clearly reveal this fact. Only inconsequential R&D spending have been invested in basic research in spite of significant technical difficulties.

As with other countries, Korea has seen a number of technical trials and errors in its machine translation research history. Especially, such grammatical theories that are suitable for foreign language but neglect the linguistic characteristics of Korean language have been adopted for use in Korea, with the result that many attempts have

been aborted. In the 1980s most research was carried out in grammatical theory. Moving into the 1990s, much concern was focused on machine translation based on corpus, but commercial products were developed separately from such research. Furthermore, the technical levels have not yet kept abreast of what is needed by general users.

Thanks to the development of the Internet, however, a desire for viewing web pages prepared in foreign language has grown, and a need for reading captions translated into Korean is increasing as opportunities of approaching foreign broadcasting programs. Such changes highlight the importance of machine translation again. Anyone can easily forecast that this field will bring forth a new market with a great potential even if not an expert in this field.

2 History of machine translation in Korea

2.1 Beginning of Machine Translation Research: Early 1980

The history of machine translation in Korea began with basic research with a small scale of research funds at 3-4 colleges including KAIST in the beginning of the 1980s. The research in machine translation during this period was started as part of a new megatrend named artificial intelligence rather than as the study in machine translation itself. As the machine translation efforts were limited to basic study and also programs designed to process Korean information were in short supply, most researches were concentrated upon peripheral, infrastructural technical fields such as Korean character codes, fonts, and Korean editor.

2.2 Optimism: 1985 - 1995

With the outset of investment in R&D for machine translation by the Korean government around 1986, researches in Japanese-Korean and English-Korean machine translation could be started. At this time, System Engineering Research Institute (SERI) of Korea and Fujitsu of Japan developed jointly Japanese-Korean machine translation system named ATLAS I/JK with the aid of Korea Japan Science Cooperation Committee. KORDIC (Korea Research & Development Information Center) used its product for JICST treatise DB service of Japan.

At the same time, with the aid of the Korean government (Ministry of Science and Technology), 4 research institutions including KAIST, SERI, ETRI, and Seoul National University and 2 enterprises including Samsung Electronics and Daewoo Telecom Ltd. developed English-Korean machine translation system named 'MATES/EK'. Separately from this research, Seoul National University began to develop English-Korean machine translation technology with the financial aid of IBM around 1988.

In 1993, 5 or 6 package type machine translation systems were released by private enterprises. 'Trainnie' and 'Word Change' are English-Korean machine translation systems: 'J Seoul/JK', 'Hangul GANA', 'Myungpoom', and 'Ogyung Paksa' belong to Japanese-Korean machine translation systems. English-Korean machine translation systems were released 1-2 years after the release of Japanese-Korean machine translation systems. But most of these products were released in such a situation that not a bit of know-how from R&D activities carried out by the research institutions was applied to them.

2.3 Disillusionment: 1995 - 1998

Moving into 1995, Enkor which was jointly developed by Seoul National University and IBM was released and POSTECH (PoHang University of Science and Technology) began R&D efforts in machine translation field. Separately from this, SERI continued on technical development of Japanese-Korean machine translation system with the aid of the Ministry of Science and Technology and finally was able to develop 'FromTo/JK', Japanese-Korean machine translation system. In addition, it developed English-Korean machine translation system named 'FromTo/EK' and Korean-English machine translation system named 'FromTo/KE' with the support from the Ministry of Information and Communications.

In spite of lots of released products and R&D activities, however, it is difficult to find such products that meet users' expectation and this situation gave rise to the discontinuance of production of most machine translation

products. As a result, enterprises involved in machine translation gave up their technical development activities and the number of researchers in academic circles was decreased greatly.

2.4 Rechallenge: Now 1999

Now, Let's look into the present situation. In the beginning of this year, an English-Korean machine translation system named 'INGUIDE' was released by LNI Co., Ltd. On the other hand, KAIST, Seoul National University, and POSTECH are carrying out projects concerning machine translation. ETRI is continuing R&D efforts in earnest jointly with 7 machine translation-related enterprises with the aid of the Ministry of Information and Communications from January, 1999 in order to put English-Korean/Korean-English machine translation systems to practical use.

3 Market Analysis

3.1 English-Korean machine translation

Presently, there are approximately 6 or 7 English-Korean machine translation systems. In the light of turnover, however, it can be said that only 1 or 2 makers' products are selling. A few products which were released 4-5 years ago show a low level of correctness in translation and draw no attention of users. Although Korea's present English-Korean machine translation market is no more than approximately US\$4 million in magnitude, it would grow rapidly as the translation correctness level heightens. If the translation correctness level would be to the satisfaction of general users in due course of time, it is expected that the market will grow to approximately US\$120 million in magnitude. Of this, 60 percent would be shared by solution market for system integration, 30 percent by package market, and the remaining 10 percent by additional value products in PC communication market.

Users of English-Korean machine translation software are higher grade school pupils, college students, and white-collar workers, and the purposes of use are varied. Young pupils are interested in translation of text in Internet and help message in game, while college students take an interest in translation of Internet texts or original English/Japanese texts. In the case of white-collar workers, they are interested in translation of fax, E-mail, or technical books. However, there is no denial that the present English-Korean machine translation systems do not meet the need of users sufficiently. On the other hand, a report says that main users may buy machine translation pro-

ducts though these seem to be of poor quality if their prices are low. In other words, users may be free to purchase the product if its present price of US\$80 drops to US\$30 or less. Also this report reveals that translation companies specializing in English-Korean translation business are not using English-Korean machine translation systems at all.

3.2 Japanese-Korean machine translation

Japanese-Korean machine translation systems began to sell from 1996 actually. The total turnover in Korea in 1997 amounted to approximately US\$2 million, but in 1998, this market did not grow greatly due to so-called 'IMF-caused depression'. At present, the Korean market in Japanese-Korean machine translation is shared by UniSoft Co., Ltd., ChangShin Co., Ltd, and Kodensa Co., Ltd. Japan. These makers are exporting approximately US\$200,000's worth of their products to Japan. Japanese-Korean machine translation market is expected to grow faster than English-Korean machine translation market because the former gives users higher satisfaction degree in translation correctness than that of the latter. If two-way machine translation, Japanese-Korean and vice versa, is developed, a market of approximately US\$12 million will emerge. The present market is shared by solution market for system integration (approx. 50%), package type (approx. 30%), and additional value market for PC communications (approx. 20%).

At the outset, Japanese-Korean machine translation products sold mainly to expert officials in research institutions out of technical curiosity. In due course of time, a movement was initiated to apply machine translation system to public relations activity or the strengthening of service beyond a level of curiosity and accordingly the system was greatly used in the business to heighten value added of the existing items. But the system is in the initial commercialization stage and is being used presently by various classes of users, and thus, the trend of purchase out of technical curiosity is on the decrease. Although it was revealed that Japanese-Korean machine translation system is not being used by translation business circles like English-Korean machine translation system is not, but there is a sign in these circles of introducing it gradually. Currently, Japanese-Korean machine translation server supporting for web navigation is released for experimental service.

3.3 Korean-English/Korean-Japanese machine translations

There is no commercial releases of machine translation softwares translating from Korean into English or Japanese. Only research institutions and colleges are per-

forming R&D activities in the prototype stage. It is expected that several Korean-foreign language machine translation systems are released in Korea before the end of 1999. The reason why it is so difficult to convert Korean into foreign languages is that Korean linguistic data forming the basis of technology for processing Korean linguistic phenomena and understanding of these phenomena are in short supply.

3.4 Machine Translation Systems available today in Korea

English-Korean machine translation systems available today in Korea are as follows:

(US\$1 = 1,200 won)

MT Systems	Company	Prices
Word Change	Jung Soft T. 82-0331-32-1592	198,000 Won
EnKor	Dream C&C T. 82-02-785-4466	297,000 Won
INGUIDE	LNI T. 82-032-785-4466	99,000 Won
NaRatMal	TechCom T. 82-053-563-0516	70,000 Won
King SeJong	MiRe Soft Center T. 82-053-563-0516	30,000 Won

Japanese-Korean machine translation systems available today in Korea are as follows:

(US\$1 = 1,200 won)

MT Systems	Company	Prices
Babel	UniSoft T. 82-02-785-4466	1,000,000 Won
i.Seoul/Jk	Dico System T. 82-02-761-2212	1,900,000 Won
i.Seoul/KJ	Dico System T. 82-032-761-2212	660,000 Won
Chosun Tong-shinsa	ChangShin Co. T. 82-02-538-8364	900,000 Won
Langfree	Daniel Tech. T. 82-02-529-8211	150,000 Won

4 R&D Activities in Korea

4.1 Conference Papers on Korean Information Processing Society

The ratio of machine translation subject on theses presented on Korean information processing field is approximately 10-15 percent. Looking into the number of theses presented to Korean Information Processing Society from 1994 to 1998, this ratio remains approximately 10-15 percent steadily. The reason can be traced from the

fact that researches have found no special way to develop the machine translation system theoretically nor actually. Another reason suggests that many researchers majoring in natural language processing would not give up machine translation.

4.2 ETRI (Electronics & Telecommunications Research Institute)

ETRI is a national research institution and has the largest think tank in machine translation field in Korea. ETRI is now carrying out a national project financed (US\$1.2 million every year) by the Ministry of Information and Communications jointly with 7 machine translation-related enterprises and some Universities including KAIST and SNU (Seoul National University). ETRI's research history of machine translation started from 1988, focused only upon basic research. The research in earnest began from 1996 with research funds amounting to approximately US\$800 thousand every year.

The project 'Development of Practical Core Technology for Korean-English/English-Korean' in execution now has an objective of turning the Korean-English/English-Korean machine translation technology, which has ended in a theoretic in-house by-product, to practical use. The purpose of this project is to inform the people, government, and sponsors of a result that can be used in machine translation field. Accordingly, what we had to do first was to choose a useful but small domain as far as possible. And we felt strongly that such a domain should have a far-reaching influence. As a result of this group discussion, we chose machine translation of English news program captions into Korean. Needless to say, we felt that we would encounter with technical difficulties such as irregistered words processing, etc. as news programs deal with various events and contents. But it was not an easy job to seek such a field having practicality and influence.

4.3 Universities

4.3.1 KAIST (Korea Advanced Institute of Science and Technology)

KAIST is developing English-Korean machine translation system named 'ALKOL (America Language to Korea Language)' relating to military information with the aid of the Air Force Headquarters. This project was undertaken in 1996 with the completion scheduled for 2000. SEO-KANG and Chonbuk Universities are involved in this project. KAIST is also developing Chinese-Korean Korean-Chinese machine translation system.

4.3.2 POSTECH (PoHANG University of Science and Technology)

POSTECH is developing Japanese-Korean machine translation system named 'COBALT J/K' for iron & steel field with the aid of POSTEEL as well as 'Collocation based Language Translator from Korean to Japanese'. In addition, it is implementing a project for Korean-Chinese/Chinese-Korean machine translation system jointly with Samsung Electronics with the aid of the government.

4.3.3 Seoul National University

Seoul National University is continuing technical development for EKTRAN English-Korean machine translation which is a subsequent version of Enkor.

4.4 Companies

In industrial circles, LNI Co., Ltd. is making a continued effort in upgrade of version for English-Korean machine translator named 'INGUIDE'. while UniSoft Co., Ltd. and ChangSin Co., Ltd are developing Japanese-Korean machine translator. Samsung Electronics Co., Ltd. is developing English-Korean/Korean-English/Japanese-Korean machine translators. Language & Computer Co., is developing English-Korean machine translator slated to be released in 1999.

4.5 Foreign Countries' Activities for Korean MT

The number of research institutions involved in studying Korean machine translation in foreign countries is greater than imagined. CMU of U.S.A., New Mexico State University, MIT Lincoln Lab. and University of Pennsylvania are employing Korean students and visiting professors to make researches in machine translation. And in Japan, companies such as Kodensa, Hitachi, and Fujitsu and universities such as Waseda, Tokyo University of Technology, and Kyoto University are said to research into Japanese-Korean machine translation.

L&H of Belgium has a special task force for machine translation in Korea to make researches in Korean-English/English-Korean machine translation systems. while Systran, Trados, and Language Partners International are said to make a study of Korean machine translation.

5 Technical Problems in the view of practical use

5.1 MT Approaches

At present, a wide spectrum of machine translation approaches is being employed all over the world. Approaches beginning with an extreme rule basis to an extreme data basis are being attempted. So is the situation in Korea. Researchers engaged in Korean information processing feel that Korean linguistic phenomena are much more difficult and complicated than other languages. A great deal of inflection or agglutinative forms are characteristics of Korean which trouble them. However, the fact is that there are no sufficient researches on how to process these characteristics successfully. Particularly, as the researches so far have been employing the rule basis research, it is also said without denial that the system development for practical use had its limit.

5.2 Corpora

There are some corpora composed of approximately 100 million words and 1 million words of tagged corpus. These corpora are not enough to support practical machine translation system technologically as yet.

5.3 Commercialization and Maintenance

To commercialize the MT technology, it is necessary to give careful consideration to the stability of system, robustness to non-grammatical texts, and duration of translation. Especially, as one can hardly imagine that technological innovation takes place in a moment to improve MT technology, the machine translation field must be researched for at least several decades. Thus, the system must be so designed that an older system can show higher translation correctness. In doing so, the system must be so developed that the existing knowledge of translation would not conflict logically with new one each time the latter is added.

5.4 General Problems

Machine translation can have good translation correctness if all the problems in processing natural language are solved. To solve such problems, it is keenly necessary to make push for joint researches by computer science majors and grammarians. But there are very few, if any, joint researches in Korea. Such miscommunications between both parties are deemed to slow down the technological development of machine translation.

6 Conclusion

As explained above, machine translation in Korea has not been matured in the aspect of market side or technology side as yet. As the need for machine translation grows, however, the investment in this field is expected to increase. To meet such a social need, a lot of investments should be made in product development together with technical development.

References

“A Plan for Development of Practical Core Technology for Korean-English/English-Korean Machine Translation.” Technical Document. ETRI, 1998.12.

Kim. T. W. “History of Machine Translation in Korea.” The 1st Machine Translation Workshop, 1999. 02.05.

Yang. S.H. “Technical Problems for Developing Practical Machine Translation System.”. The 1st Machine Translation Workshop, 1999, 02. 05.

“Natural Language Markets: Commercial Strategies,” Ovum Report. 1991

Acknowledgements

We express our heartfelt thanks to KAIST, POSTECH, Seoul National University, Samsung Electronics, LNI Co., Ltd., Language & Computer Co., Ltd, L&H, Chang-Sin Co., Ltd, and individuals who helped us prepare this report.