

THE MACHINE TRANSLATION RESEARCHES AND GOVERNMENTAL VIEW IN KOREA

Park, Changho

Systems Engineering Research Institute

39-1 Hawolgokdong Sungbukgu, Seoul

KOREA

Lee, Kisik

National Computerization Agency

80 Juksundong jongrogu, Seoul

KOREA

I . The Background and Progress of Machine Translation Research in Korea

The history of natural language processing(NLP) technologies in Korea, especially machine translation(MT), can be divided into two periods; one is until late seventies and the other is the eighties.

1. Research in NLP during the Seventies

Viewed from a broad perspective, in the seventies when we studied the basic technologies of NLP as a groundwork of MT, the focus of research was given to describing various phenomena of the Korean language in a linguistically significant way and processing the Korean characters mathematically or specific phenomena of the language logically with a computer. The theoretical linguistic description of Korean which belongs to Altaic and is an agglutinative language, mostly performed by linguists, involved such formal pattern as phonetics, phonology, morphology and syntax; and part of semantics and pragmatics was tried in tandem with them.

On the other side, a few computer scientists who wanted to establish basic technology for the NLP of Korean tried to develop a machine capable of recognizing and processing the character patterns through a formalized system of the Korean characters, Hangeui automata capable of combining and separating consonants and vowels of the language, the input-output device, a standardized system of Korean character codes, terminal and word-processor to process the Korean characters, and the like; and a few of them, with NLP techniques in mind, attempted to research the MT-oriented analysis of morphological and syntactic structures with limited samples of Korean sentences. A summary of research results obtained from the first period is as follows :

The Characteristics and Results of the Research

1) The characteristics of research

(1) In the theoretical-linguistic side, comprehensive research was done to make a linguistically significant description of Korean. Though some linguists proposed a rather simple model of Korean for the purpose of analyzing it in a mathematical and logical manner, they treated only a limited part of the language and, furthermore, they didn't aim at modelling a computational language for NLP.

(2) In the side of NLP technology, a few computer scientists researched recognition and processing of Korean character and attempted to make a computational model of Korean with an analysis of morphological and syntactic structures. However, they didn't do research for developing an applied software system based upon NLP technology like the MT system so much as for computing the Korean language in a technological/scientific way. Thus, only a limited part of the characteristics of Korean was implemented into the computer. And there were little efforts to use the research achievements for the sake of developing the NLP systems such as the MT system.

2) The results of research

(1) The achievements made during the first period include various techniques to process the Korean characters, such as automata, input/output devices with display and print functions and terminals; also we achieved basic technology for Korean word-processor capable of editing and filing in Korean and transforming the Korean characters into the Chinese ones (and the other way around). The accumulation of these techniques provided supporting environments for NLP research which became active with the beginning of the eighties.

(2) A few researchers acquired basic technologies for the processing of the Korean language through our attempt to make a computational model with an analysis of limited linguistic phenomena, which includes morphological segmentation, case classification for the interpretation of sentential structures, and such semantic elements as modality, aspect, tense, voice, etc.

With the beginning of the eighties came a new stage of NLP research, which had been only fragmentary during the seventies. The Korean Government paid much attention to promoting the information industries in the eighties, which greatly spurred the spread of computers in Korea and, at the same time, helped the technologies for processing data written in Korean ever more important. Also, as other advanced countries made practical use of NLP systems such as MT and NL-interface and emphasized the importance on NLP technology, more researchers of institutes and universities in Korea expressed a serious concern that they should make active research in processing Korean. Thus, with a more distinct purpose of developing the NLP system, the researchers began to take actions to standardize and broaden the scope of the computational model of Korean in the eighties. As one attempt, the researchers relied on various theories of processing the natural language such as Case Grammar, Phrase Structure Grammar, Transformational Grammar, GB, LFG, Montague Grammar etc.; in order to analyze the computationally implementable elements of Korean and extract linguistic data which can be testable with a computer; and with these backgrounds and aims, they began to develop a simple computational model of morphology and syntax on the basis of extracted linguistic data.

On the other hand, research in Korean character processing done during the seventies led to standardize the codes of the Korean language to be used to process it; and a great impetus was given to research in NLP when the terminal for the Korean characters was developed and the word-processor with a function of transformation between Hangeul and Chinese characters became available to ordinary people. With these environments for developing NLP technology set up, the eighties have been witnessing the flowering of NLP research in several universities and institute, which have paid particular attention to MT for the practical use and their needs.

2. Research in Machine Translation during the Eighties

MT research in Korea is being carried out mostly by several universities and national institutes, whose major concern was on analyzing morphological structure and interpreting syntactic structures. This research is based largely upon the phonological, morphological, syntactic and in part semantic descriptions of languages pre-established by professional linguists. In the early eighties, linguists did not participate in analyzing and building up a computationally implementable model, but computer engineers showed a possibility that we can develop a technique of analyzing morphological and syntactic structures of many linguistic phenomena. However, it is not until the Ministry of Science and Technology (MOST) in Korea supported the development of an MT system in consideration of its immense practical uses that full-fledged research in that area became possible.

We live in a world where various kinds of new information arise and the improvement of communication systems increases the amount of international information flow to a tremendous degree. Amid this deluge of information, the demand of translation is increasing rapidly. Under these present circumstances, Korea with its own culture based on an isolated type of language has begun to express not fragmentary but full-fledged concern about systematizing its own language.

As a point of departure, MOST decided to support research in machine-translating Korean and Japanese on a national level, since the two languages are very similar and the demand of translation from one into the other has been increasing steadily. Thus, on June, 1988, the Ministry set out to develop a prototype of the Korean-Japanese MT system with a 3-year plan. This project, led by the Systems Engineering Research Institute (SERI) of KAIST, opened a new horizon of MT research in Korea. The current research centers participated in this area in-

clude four universities (Seoul National Univ., Inha Univ., Hanyang Univ., and the Graduate School of KAIST) and a national institute (SERI/KAIST). The following summarizes the content and progresses of the project carried out by each of the above centers.

Current Research in Machine Translation at Major Research Centers

1) Seoul National University(SNU)

MT Research at this university, which grew out of its past experiences with the NLP, was started in 1983, when the researchers of the Programing Language Laboratory, Department of Computer Science, began to find out methods for English-to-Korean machine translation. Their focus of study was to make a tree-to-tree transfer from SVO-type sentences of English to SOV-type sentences of Korean. Using a method of syntactic transfer, they implemented into the computer a simple MT system which was highly restricted in scope.

And on the basis of this research experience, in 1984 they began to do research in Korean-to-English machine translation with a support from the Korea Science and Engineering Foundation (KOSEF). With about 150 sample sentences selected from the middle school textbook of Korean history, the researchers at SNU attempted to make an MT system equipped with the morphology analyzer, the parser for the Korean, the translation grammar, and the machine dictionary. For the analysis of sentential structures, they relied on Transformational Generative Grammar. To realize it, they adopted Marcus deterministic parsing and C-language. This translation system has about 500 lexical units, which are used at each phase of processing - the morphological level, the syntactic level, and partial semantic level, etc.

Since July, 1986, this university has also been developing an English-to-Korean MT system called KSHALT in collaboration with IBM. It aims at translating computer manuals by means of PEG(PL/NLP English Grammar), a parser of English provided by IBM Watson Research Center. Most of its concern is on the transfer from English to Korean and the generation of Korean sentence.

2) Inha University

The Department of Electronic Engineering at this university began to do research in the morphological analysis of Korean and a parser for the Korean in 1978; and its research in bidirectional machine translation, which began in 1983, was led to developing a prototype system in 1985. This system, mainly for translation between Korean and Japanese, runs on IBM/AT; and the size of dictionary is about 40,000 lexical units. What is notable in this system is that it adopts a method of flexible combination network composed of three patterns: P-pattern, V-pattern, and Q-pattern.

P-pattern: a case pattern composed of pairs of semantic primitive and case

V-pattern: a pattern of the elements obligatorily governed by verbs

Q-pattern: a pattern of the elements optionally governed by verbs

These three patterns combine each other in a flexible manner to form a pattern network generating an infinite number of structures. The pattern network analyzes the syntactic-semantic structures and modification relations of sentences simultaneously; and this network performs interpretation and generation into target language bidirectionally. Currently, researchers at this university are trying to equip the system with inference functions to solve syntactic and semantic ambiguities and other semantic problems, while showing serious concern about how to summarize Korean texts and how to upgrade performance of the system. On the other hand, they are testing the efficiency of the system with regard to Korean-to-English translation.

3) Hanyang University

For the sake of machine translation, this university produced an English parser as a result of research in NLP technologies, and implemented PAMT to parse sentences for general use. Also, since 1986, it has been working for a Japanese-to-Korean MT project in collaboration with Waseda University of Japan. PAMT is a device of syntactic analysis for machine translation which was developed by improving Tomita's algorithm capable of interpreting context-free grammar rules with an extended LR parsing method.

Researchers at this university relied on Lexical-Functional Grammar and have developed programs to compile grammatical rules and their dictionary, and generate well-formed sentences. PAMT, with the transfer phase built up independently of the generation phase, can be applied in various ways, depending on how the dictionary is compiled and how the semantic informations are represented. Currently, this university is making efforts to apply it to the discourse level as well as MT. On the other hand, the researchers are working hard to develop a prototype system in their joint research with Waseda University which aims at translating Japanese videotex into Korean.

4) The Graduate School of KAIST

Since they set out to develop a parser for the Korean language in 1982, researchers at the Graduate School of KAIST have tried to make computational linguistic models of Korean on the basis of Case Grammar, Montague Grammar and Lexical-Functional Grammar mainly. Recently, they are also working to develop techniques of interpreting Korean in recourse to Generalized Phrase Structure Grammar. Also, in 1985 they reinforced their research in the MT system and set out to work for Korean-to-Korean machine translation using an inter-lingua.

On the other hand, in 1986 they organized a cooperative team with the NEC of Japan in order to develop an MT system which is capable of multi-lingual translations including Korean-Japanese and Korean-English. It is known that this system will be based upon the pivot approach. For this purpose, the researchers are currently studying the analysis and generation of Korean.

5) The SERI of KAIST

Systems Engineering Research Institute (SERI)/KAIST has carried out two MT projects: One is a national project aiming at Korean-Japanese (and vice versa) translation, and the other is a joint research with France in English-to-Korean and French-to-Korean translation.

MOST, which chose MT research as a national task on June, 1983, commissioned SERI to carry it out for about 3 years. This was the first national project of MT supported by the Korean Government. The objective was to develop a MT system between Korean and Japanese. The reason for choosing these two particular languages as the first step is that they share many linguistic characteristics and the demand of translation in them has been increasing steadily.

The study on Korean-to-Japanese translation was intended to be a means of improving and expanding the established NLP research for Korean. The system is based upon the syntactic transfer; and it makes a flexible translation with the translation rules in packets of the Grammar Writing System and the dictionary rules describing linguistic knowledge into the dictionary. Also, the Japanese-to-Korean translation system, set forth as a part of national project, was later chosen as a joint project of Korea and Japan with the help of the economics associations in both countries. And it was developed in collaboration with Fujitsu in Japan.

As a result of the national project, SERI individually succeeded in developing the KANT/I system which aimed at MT between Korean and Japanese. For Korean-to-Japanese translation, it operates on the 16 bit UNIX machine. And for the Japanese-to-Korean translation system, SERI has made more efforts to develop it with Fujitsu as a cooperative project between both nations. Through this project, Fujitsu developed analysis modules of Japanese and SERI took the responsibility of developing the transfer module of Japanese-to-Korean and generation module of Korean. In 1985, both organizations developed a prototype system which aimed at translating Japanese to Korean. After improving its functions/efficiency and evaluating its productivity, they are currently trying to manufacture it with the FACOM/M series.

Research in English-to-Korean and French-to-Korean machine translation, carried out collaboratively with GETA/CNRS in France, was set up on November, 1986 with a support from MOST. Its first objective is to make an English-to-Korean translation system. Currently, both GETA and SERI are in feasibility study of English-to-Korean translation by using the ARIANE system of GETA. And it is expected that the full-fledged study will begin in late 1987 as researchers of SERI/KAIST are sent to GETA.

II. Government's Viewpoint and Assistance for the Machine Translation

The Ministry of Science and Technology (MOST) is the governmental body chiefly responsible for assisting MT research in Korea. The MOST assistance program dating back to 1983 when the Korean-Japanese MT project, the first of its kind in Korea, was launched today covers Korean-English MT research as well. In Korea, the translation between Korean and English is in the highest demand. The primary reasons for the selection of Korean-Japanese MT research as the initial target of government assistance were that there was relatively high demand for translation between the two languages and that the research didn't involve much risk of failure due to the linguistic affinity between the two languages.

The MOST is also assisting the development and training of research manpower at universities by aiding their small scale NLP research projects through the KOSEF under its umbrella.

Another area of interest for the MOST is the development of Korean NLP technology by drawing on Artificial Intelligence(AI) technology. In order to attain this goal, the Korean NLP research is incorporated into the national projects for the development of AI technology and universities and research institutes are commissioned to conduct it.

The motives for the continued assistance of MOST for MT research in Korea are three-fold.

- (1) MT is considered inevitable to respond to the growing public call for time and energy saving translation systems, and the acquisition of workable Korean NLP technology is viewed as urgent in connection with the development of next-generation computer software.
- (2) This requires the improvement of the qualifications of research personnel at universities and research institutes and the upgrading of their research levels.
- (3) The Korean-Japanese MT project has brought about positive results and, thus, has made it easier to solicit government assistance for MT research projects.

1. Government's Viewpoint

The MOST evaluates that the development of the Japanese-to-Korean MT system, jointly conducted with Fujitsu of Japan has been a successful project. The system, which connects the user's peripheral with a large host computer, is now in the stage of being tested for commercial application. At the same time, the ministry recognizes the importance of acquiring Korean NLP technology capable of analytical and intelligent processing of the Korean language. To achieve this goal, the ministry is striving to foster a more efficient cooperative relationship among universities and research institutes and work out comprehensive and long-term assistance programs for their research projects.

The fact, however, remains that a few decision makers at the MOST, even though they admit the necessity of MT, are skeptical of its successful commercial application. This skepticism stems from the low job performance due to unsatisfactory translation quality, low cost-efficiency and limited usage of the already developed MT system. Consequently, they have a negative attitude toward the expansion of government assistance for MT research.

As a matter of fact, the first Japanese-to-Korean MT system, running on a micro-computer and marketed last year, has not received a very positive response from the public because it failed to overcome the historic shortcomings of MT, namely, poor quality and low productivity. This has sent a warning signal to the companies which were either already engaged or interested in MT research. As a result, today they became more cautious about participating in the research.

Despite such skepticism and negative views, however, the MOST remains firmly committed to rendering continued assistance for MT research as a groundwork for the ultimate R/D of NLP technology. Along with this, the ministry is said to be considering entrusting the AI Research Association (established on Dec. 1085) under the Korean Information Science Society (KISS) with the task of spurring the sophistication and expansion of MT system research to ultimately in-

clude the development of intelligent NLP technology capable of *understanding* the natural language.

2. Prospects for Government Assistance

The MOST has yet to announce any concrete plans for aiding MT research projects in the long term. But the past history of MOST assistance for MT projects and the on-going national research projects indicate that the MOST will focus on two particular areas.

First, despite technological limitations, the ministry is expected to continue its assistance for MT projects designed to develop practical translation systems for particular language pairs where demand is relatively high.

Second, it is believed that the MOST will vigorously push forward with Korean NLP research as part of an effort to develop intelligent next-generation computer technology. In doing so, it will draw on the results of artificial intelligence researches.

The MOST assistance in the first area will be implemented on a short-term, case-by-case basis. By contrast, in the second area the ministry will act under long-term plans. The plans will call for building a cooperative research structure linking universities and national research institutes and, if necessary, foreign research organizations. Examination of the on-going research projects sponsored by the MOST enables us to more clearly forecast the future direction of Government assistance for MT research.

(1) The MOST, in addition to the Korean-Japanese MT project, has commissioned the SERI/KAIST to conduct a joint research project with GETA/CNRS of France to develop English-to-Korean and French-to-Korean MT systems.

(2) The MOST now has two national projects underway aimed at the R/D of Korean NLP technology based upon artificial intelligence.

In the first case, the ministry arrived at the conclusion that the development of an MT system can be more effectively done if the project is jointly conducted with the nation whose language is involved. Accordingly, at the Korean-French Science Ministers' meeting in Sept. 1985, Korea and France reached an agreement to conduct joint MT research. Under that agreement, SERI of Korea and GETA of France were chosen as research partners to first develop English-to-Korean MT system. The MOST will support the joint research project for the coming three years. In the first two years, the objective will be to develop a prototype system for research purpose. Then, a feasibility study will be conducted as to the commercial applicability of the system. Based on the results of the study, related industries in both countries will be called in for joint production of the system for commercial use.

The two national projects on the development of Korean NLP technology are being carried out by universities and research institutes under the supervision of the AI Research Association at the KISS.

The first project, which was launched early this year, is led by universities and aims at developing Korean NLP technology for MT application. The second project will start in 1988. The objective of the project to be led by research institutes is to develop Korean interface or front-end technology. The main thrusts of the project are development of natural language analyzing technology and related tools with a larger scope of application, and researching natural language understanding technology capable of intelligent Korean NLP.

Therefore, it is expected that AI techniques such as knowledge representation and inference mechanism, etc will be more actively introduced to facilitate the research on NLP technology. To overcome the problems experienced by traditional MT researches, it is expected that NLP research in Korea will shift its focus from the development based upon the syntactic analysis technology to that based upon semantic and partially pragmatic analysis.

In this sense, the MOST has recently decided to revise its assistance policy in such a way that will propel coordination among research institutes and it is now gathering opinions from related academic institutions and research organizations. In the academic community, the AI

Research Association has recently sponsored a NLP workshop with the participation of members of the Korean Linguistic Society and the Cognitive Science Society, where they explored the possibility of a joint participation in NLP research.

As mentioned above, the R/D Projects of MT which have constituted the core of NLP research in Korea have been principally aimed at developing Korean-English and Korean-Japanese translation systems. The systems, however, are still far away from being commercialized. This is especially true in the event Korean is the source language because of the lack of research work to analyze the linguistic phenomena and properties of the Korean language.

In this context, the MOST will strengthen government assistance by realigning the research system and increasing its research assisting budget in order to spur the early acquisition of NLP technology and the development of practical MT technology.

- Under the reoriented research system, universities will lead basic research on NLP technology including MT, and research institutes will take the primary responsibility for integrating all available technologies and developing a prototype system as an initial step for the development of the targeted system in the long run. The companies will be allowed to join the project at the stage of practicalizing the prototype system in the form of financing partial budget for production of the commercial system.
- As for budgetary assistance, it is clear that the MOST will increase its financial support for MT research gradually based upon its evaluation of the importance of each project. The financial aid that the MOST has extended for MT research since 1983 upto now reaches nearly one million dollars.

In addition, the Ministry of Communication (MOC) is planning to include the utilization of MT system technology in its master plans for the development of information and communication technology. It is forecast that in the long run the MOST and the MOC will move in the direction of joint participation in MT research.