[*Progress in Machine Translation*, ed. Sergei Nirenburg. Amsterdam: IOS Press, 1993.]

Machine Translation: The Japanese Experience

• MAKOTO NAGAO

1 Introduction

Practical machine translation systems are operated more broadly in Japan than anywhere else in the world. There are several Japanese-to-English and English-to-Japanese MT systems available on workstations and portable WSs). Japanese researchers and developers know well that natural languages are too complicated for an elegant and simple linguistic theory to describe all linguistic phenomena. They have striven for a balance of basic linguistic research and *ad hoc* treatment of more intricate areas of language use. They have also developed user-friendly interfaces for pre-editing and post-editing of texts, which are central in the present-day MT systems. With these efforts many users of MT systems have managed to reduce translation costs by 30 ~ 50% as well as to speed up the translation production process. The huge investments by manufacturers of MT systems cannot, of course, be compensated through the sales of a few hundred or a few thousand MT systems. However, top management of these corporations realize that R&D for MT systems will be a basis for the natural language processing technology which is definitely a key technology in the next century.

2 History of MT Research and Development in Japan

A partial list of major machine translation R&D activities in Japan is given below.

1955 MT research begins at Kyushuu University. Designed and completed a dedicated computer for multilingual translation (Japanese, English, French and German) in 1959 (Tamati 1985).

1955 research on a phonetic typewriter begins at Kyoto University. First demonstration was given 1960 (Sakai and Inoue 1960).

1956 Electrotechnical Laboratories started research toward an English-Japanese MT system. Demonstration at the first IFIP Congress in Paris in 1959.

1962 MT research begins at Kyoto University.

1964 Random generation of sentences based on semantic dependency of components in a phrase and on application of probabilistic rules (Nagao 1965). A thesaurus-type semantic marker system is introduced.

1973 PLATON: a programming language for grammar writing and parsing is developed (Nagao and Tsujii 1974).

1974 Case grammar gains popularity for Japanese sentence analysis. Contextual analysis of Japanese sentences (anaphora and ellipsis resolution model) studied (Nagao and Tsujii 1976).

1976 An electronic dictionary system (Japanese, English-Japanese) is developed (Nagao et al. 1978) and used for various natural language processing tasks as well as for the extraction of thesauri and other semantic information.

1978 A commercial Japanese word processor (of the Kana-to-Kanji conversion type) appeared in the market. In widespread use from 1980. Essential part of the word processor is a Kana-to-Kanji conversion dictionary.

1979 TITRAN: title translation system of scientific and technical papers from English into Japanese was developed and used for the translation of information in the INSPEC database (Nagao and Tsujii 1979). TITRAN J-E and TITRAN J-F were demonstrated in 1980 and 1983 respectively.

1980 An MT system for translating computer manuals from Japanese into English was first demonstrated.

1982 The Mu machine translation project sponsored by Japanese Government began (Nagao 1983a). It aimed at the translation of abstracts of scientific and technical papers between Japanese and English. The Japanese-to-English system was demonstrated in 1983, and the English-to-Japanese system was demonstrated in 1986. The latter included a dictionary of 70,000 entries. The project gave a strong impetus to machine translation activities in Japan. It was completed in four years and cost 650 x (10^6) Yen (nearly US$4,000,000). The MU-II project was carried out in 1986-1990 at the Japan Information Center of Science and Technology for the development of an operational system. Since January 1991 the center regularly performs Japanese-to-English translation of abstracts accumulated in its database. Translation cost has been reduced by fifty per cent compared with human translation cost. The system has a dictionary of 500,000 entries. The size of the dictionary is an important factor in the performance of the system.

1983 A controlled language was proposed, and a writing aid for this controlled language was developed which included a unique parser was demonstrated (Nagao 1983).

1984 Commercial systems (Fujitsu, etc.) are announced. Over ten systems have been marketed until present.

3 The Japanese Government Attitude to MT

Japanese Government, particularly the Ministry of International Trade and Industry (MITI), the Ministry of Post and Telecommunication (MPT), and the Science and Technology Agency (STA) have been supporting varieties of projects in information processing and distribution.

1972 The Pattern Information Processing System Project (MITI), ten years. Primarily, recognition of speech, characters, images and language.

1982 The Mu Project (STA), four years. Machine translation of abstracts between Japanese and English.

1982 The Fifth Generation Computer Project (MITI), ten years.

1986 The Mu-II Project (STA), five years.

1986 The Electronic Dictionary Research Project (MITI), nine years. Development of Japanese, English and general dictionaries.

1986 The ATR Interpreting Telephony Research Institute (MPT), fifteen years. Speech translation research between Japanese and English

1987 The CICC Multilingual MT Project (MITI), 6 ~ 8 years. A cooperative project between Japan, China, Malaysia, Thailand and Indonesia.

1987 The MT Summit at Hakone (MITI).

1988 The Friend 21 Project (MITI), six years. Research into user-friendly interfaces.

At present, additional government agencies in Japan have expressed interest in machine translation.

- The Patent Office of Japan is interested in MT.
- Mm is using commercial E-J MT systems for internal use.
- JICST is operating a J-E MT systems since January 1991 to translate abstracts from Japanese into English.
  Mm seems to be interested in world-wide cooperation in the development of multilingual MT systems. It will include the standard specification of interlingua and the construction of a multilingual dictionary.

4 The Attitudes of MT Developers

1. Successful sales of Japanese word processors (the market is currently estimated at $2 billion annually) helped the business community to recognize the importance of natural language processing and machine translation.
2. the potential annual MT market in Japan is estimated at several billion dollars or $10^{11} \sim 10^{12}$ words per year (JEIDA 1989).

3. At beginning of the 21st century friendly human interface technology will be a central issue. Speech and natural language technologies are to be established rapidly, and applications to realize office automation must be developed.

4. Japanese manufactures have the following views for the R&D of MT. Machine translation is one of the typical applications in the future market. Machine translation includes varieties of basic natural language technologies, and therefore must be developed.

5. These technologies will be applicable to the development of intelligent information/text retrieval systems, complex hypertext/hypermedia systems and flexible human interfaces for a variety of applications.

6. Manufacturers have already invested huge amounts of money and manpower, and therefore cannot withdraw from MT easily. A manufacturer cannot stop R&D of MT unless other competing companies do so at the same time. Dropping out of this competition spells defeat in the future big markets of the information society.

7. Several manufacturers have already passed the break-even point counted on the basis of annual income and expenditures (not counting the development cost) and are posting small profits, so that there is a wide belief that the second ALPAC report will never be written (JEIDA 1989).

8. Manufacturers know well that a single linguistic theory cannot lead to a good MT system. They realize that a huge amount of language phenomena must be processed in an *ad hoc* manner. This requires enormous efforts for the improvement of grammars and dictionaries. R&D of tools for MT research must be supported to help with the grammar and dictionary acquisition effort.

9. Dictionaries in commercial MT system are proprietary. Manufacturers don't like to disclose their details, nor share them with others. Dictionary acquisition is very expensive, and the duplication is not desirable. But manufacturers invest money in their own dictionaries, however redundant they may be. Japanese government project in electronic dictionary research aims at resolving this difficulty.

10. Computer technologies advance rapidly, and in several years personal computers will become powerful enough to be equipped with complicated machine translation systems relying on a fairly big dictionary. Once this happens, the new tasks to which machine translation becomes applicable will accelerate the improvement of MT systems.

11. All engineering products display their full ability when they are used within the limited range of the design. The same is true for machine translation systems. There is a movement to design a controlled language which can be analyzable by average machine translation systems. This idea is good not only for machines, but also for human beings, because a controlled language decreases ambiguous expressions, and help human beings to understand the contents clearly and uniquely without misunderstanding.

12. Machine aids for human translation, or machine-aided human translation systems do not contribute very much to the real advancement of natural language technologies, and therefore Japanese manufacturers have little interest in them.

5 Attitude to MT in the User Community

1. In the past two extreme views on MT were prevalent — some people had excessive expectations for MT, others denied even the possibility of MT.

2. The number of people who have interest in MT and have a certain knowledge of MT is growing. They want to utilize systems wisely, within the systems' capabilities. There is, therefore, hope for the society to accept MT systems and to use them properly.

3. Users of MT systems must have proper linguistic knowledge about sentence categories which are acceptable by MT and which are not. Some users complain about the bad quality of MT or about the failure of translation without checking how bad or ungrammatical the original sentences they intend to translate are.

4. Many Japanese users or potential users are already wise enough to use present-day MT systems in a cost/effective way. The number of such users is increasing, and they know that they have to prepare some dictionary terms and have to participate in the improvement of their systems in cooperation with the manufacturers.

5. End users (or readers) of translated texts in Japan are generally generous or tolerant to the quality of translation. They don't demand the best quality. They accept the translations when they are understandable after a moderate effort.

6. There are many Japanese people who can read English but who want quick scanning of headlines or abstracts in Japanese. For this class of users, raw machine translation results are sufficient. They will read the original English texts when necessary.

7. There is a large volume of documents to be translated either into English or into Japanese at present, and the volume is increasing year by year. Almost none of these documents require literary translation, but all require exact, unambiguous translation. Many documents include difficult technical terms from the sciences, economics, politics, etc. Therefore many translators feel that they cannot avoid machine translation if they want to achieve better translation speed and cost performance. There are several training schools for post-editing of MT in Japan. Younger translators hold rather positive attitudes toward MT.

6 Researchers' Attitude to MT

1. Japanese researchers in natural language analysis and machine translation direct their basic research toward analyzing real texts. Sentences which are discussed by theoretical linguists are generally too short. They may be interesting from the point of linguistic theory, but they are very different from sentences of real texts. The latter contain difficulties which linguists do not usually discuss, and which are clearly revealed by the application of MT programs.

2. Simple and elegant linguistic theories are irrelevant and helpless when applied to real text. Several frameworks to cope with the difficulty of this kind were proposed and implemented in computer programs in Japan. They may not look beautiful and theoretical, but they constitute good basic research in computational linguistics.

3. Although there is very little cooperation in research and development between universities and industry, people in industry constantly watch the activities and results at universities and use the better research results in their work. As a result, there is no serious problem of technology transfer in Japan.

7 Future of MT

1. Machine translation software will be installable on a word processor or on a laptop personal computer in the near future, and machine translation will become available everywhere, at home, in the office and at school. In such a situation a system will adapt itself quickly to individual users through the application of the learning mechanism in the system.

2. Another direction is the network access to a big MT system from a home computer. Users in this category may not use MT systems regularly, and therefore special care must be taken to develop user-friendly interfaces.

3. Present-day commercial MT systems have a definite limitation in the translation quality because they are based on the compositionality principle. Anaphora and ellipsis cannot be handled well because the systems do not see intersentential relations. A significant amount of advanced research in this area must be performed.

4. Multilingual machine translation systems must be developed not only among European languages but also among Asian languages and languages world over.

5. Speech translation research must be conducted. At ATR Interpreting Telephony Research Institute in Japan, a research project is going on since 1986 for Japanese and English. It is highly recommended that other countries start similar research projects.

6. International cooperation is essential in machine translation research and development. Dictionaries must be exchanged, and contrastive studies of languages are to be promoted as cooperative research among different countries. Such activities must be supported continuously for a long term, for example, for ten or fifteen years, because language is a very difficult object to study. Agreements must be reached at the governmental level for such international cooperative research (NRC 1990).

References

JEIDA. 1989. A Japanese view of machine translation in the light of the considerations and recommendations reported by ALPAC, U.S.A., July

Nagao, M. 1965. Sentence generation by semantic concordance. Presenetd at *Int.Conf. on Computational Lingusitics*, New York.

Nagao, M. and J. Tsujii 1974. PLATON: a programming language for natural language processing. *Journal of Information Processing Society of Japan*, **15**(9). (also in American Jour. CL, Vol.13, No.1, Microfiche 37, 1976).

Nagao, M. and J. Tsujii 1976. Ananlysis of Japanese sentences using semantic and contextual information. *American Jour. CL, Microfiche 41*. (alos in Proc. COLING 76).

Nagao, M., J. Tsujii, A. Yamagami and S. Tatebe. 1978. Japanese dictionary in computer, and an automatic segmentation of Japanese sentences. *Journal of the Information Processing Society of Japan*, **19**(6). (also in Proc. COLING 80).

Nagao, M. and J. Tsujii 1979. A machine translation system from English into Japanese for titles of scientific papers. Presented at *2nd Int. Seminar on MT, Moscow*, 11 (also in Proc. COLING '82)

Nagao 1983. A trial of controlled language. Presented at *IPSI Natural Language Processing Symposium, 6.*

NRC. 1990. National Research Council. *Japanese to English machine translation*. National Academy Press.

Sakai, T. and S. Inoue. 1960. New instruments and methods for speech analysis. *J. Acoustical Society oif America*, **32**(4).

Tamatai, T. 1985. A historical overview of machine translation. *Journal of the Information Processing Society of Japan,* **26**(10)