[*Translating and the computer 13: The Theory and Practice of Machine Translation – A Marriage of Convenience?*
Papers presented at a conference … 28-29 November 1991 (London: Aslib, 1992)]

CURRENT PRACTICAL MACHINE TRANSLATION SYSTEMS IN JAPAN AND FUTURE DIRECTIONS

Makoto Nagao

Kyoto University

INTRODUCTION

Just a year ago in November 1990, an American study team came to Japan to see the state of the art of machine translation (MT). They visited more than twenty manufacturers and users of MT during a week. The team was composed of ten people headed by Prof J Carbonell, Carnegie Mellon University. The study was promoted by the Department of Commerce, National Science Foundation and DARPA. Their report will come out soon, but clearly indicated that the United States of American has come back to this very futuristic, but interesting area after a long sleep since the ALPAC report in 1966.

MACHINE TRANSLATION IN JAPAN

Japan has been making constant efforts for research and development of machine translation systems. Major computer companies have been involved in this area, because they are not only interested in commercial products of machine translation but also they know that natural language processing technology is a key in the coming information society.

There are many machine translation systems commercially available now. Typical systems are listed in Figure 1. There are some other small systems on personal computers, and also there are many developments going on in many companies. Some companies claim that they sold more than one thousand, several hundreds and so on. But some other companies are rather careful in selling a lot of systems because they are anxious about heavy load of after-service or maintenance which is required constantly. Users are always complaining about the quality of translation and they ask the producers to improve the systems. This after-service is a key factor to enlarge the market share for the producers.

There are many systems which are once used but abandoned because of not good after-service. These are generally the systems from earlier years of commercialization of machine translation. Recent systems have been improved significantly, and the producers have established good after-service networks, and therefore their systems are used constantly. I guess there are one to two thousand systems actually used for translation on a daily basis all over Japan.

There is a significant change in hardware and software of machine translation systems (see Figure 2). In the middle of the 1980s all the systems were on main frames. They were expensive and inconvenient to use. Recently all the systems have been realized on workstations. They are relatively cheap, but sufficiently powerful, and very convenient for translators to use. There is a gradual shift, however, from workstations to personal computers because personal computers have become powerful, workstations have become small and portable and there is no clear distinction between workstations and personal computers. This shift will become significant in another one or two years. Then machine translation systems

Organization	Language Pairs	Status
ATR	J-E, E-J	speech translation, research
Catena/NHK	E-J	
CICC	J, C, Thai, Malay, Indonesian (all pairs)	research
CSK	J-E, E-J	
Fujitsu	J-E, E-J, J-G, J-C, J-F, J-K	some pairs are prototypes
Hitachi	J-E, E-J	
Japan IBM	J-E, E-J, E-C, E-K	some pairs are prototypes domestic use
JICST	J-E	specialized for abstracts domestic use
Matsushita	J-E	
NEC	J-E, E-J	
NTT	J-E	development
Oki	J-E, E-J	
Ricoh	J-E, E-J	
Sanyo	J-E	
Sharp	J-E, E-J	
Toshiba	J-E, E-J	

- Some other systems on PC's.
- There are several research systems at some universities.
- J: Japanese, E: English, C: Chinese, G: German, F: French, K: Korean.

Figure 1 Japanese machine translation systems

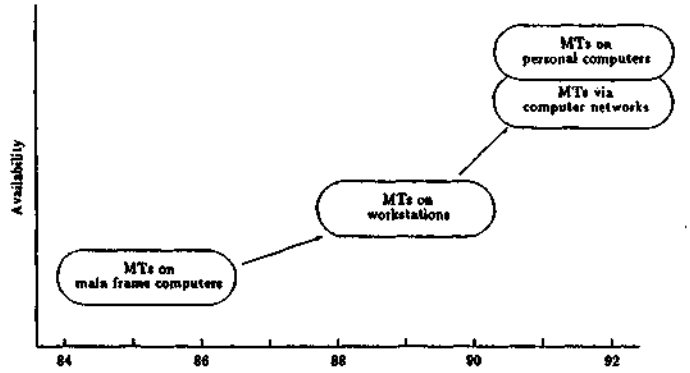

Figure 2 Development of machine translation systems

will come into home, and the users will increase greatly. They will make their own effort to improve the systems to achieve better translation quality, and this will make the producers more and more advanced in machine translation systems because the producers get a lot of translation know-how from users.

Another direction of machine translation usage is via computer networks. There are already a few machine translation services available in Japan via computer networks. In this situation machine translation systems are on powerful computers and varieties of large dictionaries are available. This mode of usage will be profitable for casual users of machine translation. However the terminal computers for users must be equipped with good editing software to edit the translated results.

Translation in Japan either by human or by machine is mainly between English and Japanese. Figure 3 shows that there are some other languages from which the translation is done into Japanese, but the translation of the Japanese documents are almost all into English. These statistics are only from the translation firms. There is a lot of translation done inside companies where the translation is almost all between Japanese and English.

Figure 4 shows the kind of texts translated by machine. We can see the dominant use of machine translation systems is for technical manuals, product specifications and technical reports used inside a company. We have no statistics about what percentage machine translation contributed to the whole volume of translation in Japan.

A typical translation process by computer is the following:

1. input by OCR, floppy disk or via computer network
2. pre-editing
3. machine translation

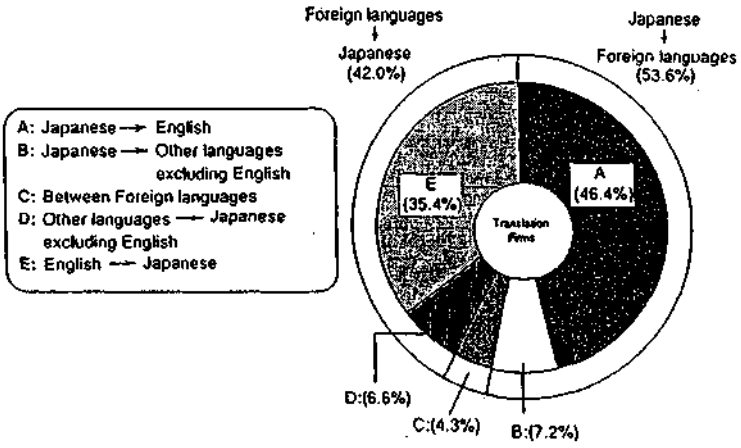

Figure 3 Translation language pairs in Japan (human and machine translation)

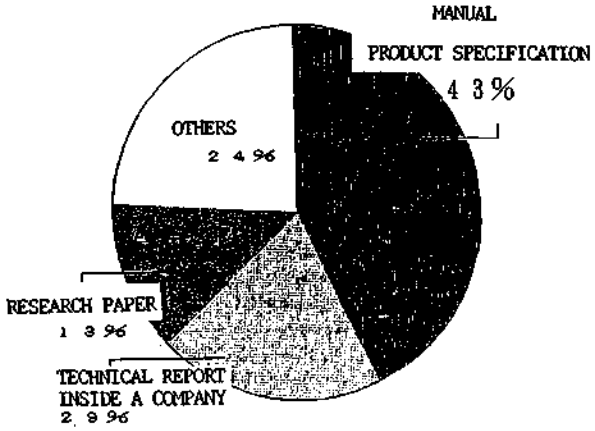

Figure 4 Kind of documents translated by machine (Japanese to English)

4. post-editing
5. check by native speaker
6. print-out.

There are discussions about what kind of pre-editing is good for machine translation systems. It heavily depends on the kind of documents and also on the system used. A typical specification for pre-editing technical manuals in Japanese is:

1. change declarative sentences into imperative sentences
2. change noun phrases into declarative sentences by adding proper predicates
3. supply subject and/or object in case of omission
4. use Chinese character expressions whenever possible
5. eliminate redundant (post-positional) expressions and make a sentence as simple as possible
6. divide a long sentence into several short sentences. This is particularly important when a Japanese sentence has more than eighty characters or an English sentence is composed of more than thirty words.

Translation quality is heavily influenced by the nature of input sentences. Some documents need heavy pre-editing and others are not. There is a movement in companies to write documents clearly so that not only humans can understand without ambiguity but also machines can translate successfully. However there is no such specification established yet. Figure 5 shows somewhat symbolically the effectiveness of editing, and the improvement of machine translation systems.

Translation quality is the most important factor in machine translation. How to measure the quality has been discussed and tried since the ALPAC report. Accuracy (or fidelity) to the original text, and the intelligibility (or readability) are the major factors in the evaluation of translation quality. However, the acceptability of machine translation systems is largely dependent on what quality users expect from machine translation systems. Generally speaking Japanese people are more patient or generous about the poor quality of translation than Europeans and Americans. Factors other than accuracy and intelligibility are sometimes more important, such as turnaround time, terminological consistency through a whole document, high throughput, and so on. These factors are very often superior by machines than by human translation. Some of the existing systems in Japan are already good enough if used as productivity tools by professional translators and also if used for quick information scanning.

However, direct machine translation output is often incomprehensible, and pre-editing, post-editing, or both are necessary. Therefore developers are focusing their efforts on the development of support tools for machine translation. Support tools include not only the software for pre-editing and post-editing, but also dictionary editors, online diagnostic tools, grammar checkers, unknown word detection software, and so on. Word processors and machine translation systems are being combined.

There are two major categories for document translation. One is for assimilation to monitor foreign technology. For this purpose raw machine translation output is generally enough. The translation is from many foreign languages into Japanese. Another use is for dissemination of information. A typical application is product manuals for overseas customers. Translation quality must be high, and so when machine translation is used, post-editing is inevitable, and pre-editing is sometimes used as well. The translation is from Japanese to foreign languages, mainly English.

Cost for machine translation is not disclosed from translation companies. They do not like to say that this document is translated by machine translation system because they charge the same price for machine translation output. They claim that they do sufficient post-editing to

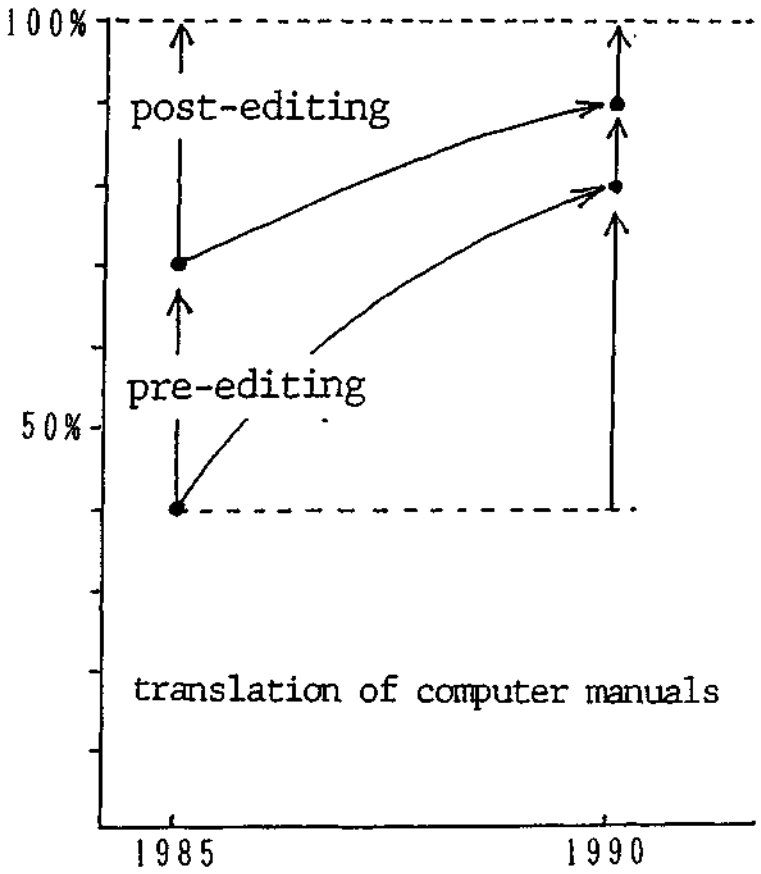

Figure 5 Effectiveness of editing

the machine translation output, and the quality is the same as the human translation output, and so there is no need of the distinction.

An example shows that raw machine translation output costs 180 yen/page (one page is about 200 English words) for English into Japanese, and 360 yen/page for Japanese into English. These correspond roughly to 7% and 10% of human translation costs respectively. Another example shows that the cost of Japanese to English translation output after post-editing is between US $40 to $70, which corresponds roughly to 70% of human translation. Users of machine translation systems expect that the advantages of machine translation are 1) consistent use of technical terminology, 2) less time spent on dictionary consultation, 3) quality more even than with human translation, 4) improved turnaround time, 5) reduced cost, and 6) integration with publishing systems. These are particularly true for the translation of large volumes of text.

Users of machine translation systems have keen interests not only in translation quality, but also in the integration with automatic input/output systems, user-friendly interface for customization of systems and pre-/post-editing. They are also interested in cheaper and more convenient systems such as personal computer systems with the same or more powerful translation software.

Users complaints on machine translation systems are poor quality, high cost, slow turnaround time, and operational difficulty. Among these poor quality is the most serious problem, and others are far less.

Government Programmes

The Japanese government has been constantly supporting machine translation research and development. The Science and Technology Agency supported the Mu project during 1982-86. The project was to develop prototype machine translation systems (Japanese to English, and English to Japanese) which were dedicated to the translation of abstracts of scientific and technical papers. The Japanese to English machine translation prototype was converted to a practical system at the Japan Information Center of Science and Technology, and the system has been in daily operation since December 1990. A multi-lingual machine translation project is going on by the support of the Ministry of International Trade and Industry (MITI). It is an international cooperative research and development between Japan and four Asian countries, namely China, Thailand, Malaysia, and Indonesia. MITI is also supporting an electronic dictionary research project, which is conducted by the consortium named as EDR by eight major electronic companies in Japan. It aims at the construction of machine usable dictionaries of Japanese and English, and also bi-lingual dictionary between Japanese and English via neutral concept dictionary. The dictionary size is 200,000 general words and 100,000 terminological words in the field of computer science. The Ministry of Post and Telecommunication is supporting a very ambitious project of speech dialogue translation between Japanese and English. ATR Interpreting Telephony Research Institute is the responsible organization for this fifteen year project.

Besides the steady step improvement of the present-day machine translation there are several ambitious attempts in machine translation research. One is the so-called probabilistic approach. It calculates word order statistics for quite a large volume of bi-lingual texts, and establishes the stable translation phrase pairs. It may work well for such language pairs as English and French, but is quite doubtful for English and Japanese, because these two languages have quite different language structures.

Another new hopeful approach is example-based machine translation. This was proposed by the author in 1984, but attention was not given to this until recently. The idea is that the translation is guided by example translation. This is in a sense similar to what a human does in translation. We prepare lots of translation pairs and also a word thesaurus. When a sentence is given for translation, we look for an example phrase which matches best to a fragment of the given sentence. This best match is calculated by utilizing thesaurus information, that is, a kind of distance between words. This operation is performed until a set of example phrases covers the whole input sentence. Then we convert a matched fragment into a fragment of the target language by seeing the translation of the example phrase. This operation is done to all the matched fragments and then we compose these translation fragments into a sentence. This example-based translation gives relatively high quality translation compared to the present-day ordinary machine translation mechanisms which depend on semantic primitives for the selection of proper translation words.

Speech to speech translation researches are going on in Japan. ATR Interpreting Telephony Research Institute started the research five years ago and then research and development sections of some private companies followed. ATR's objective is to translate Japanese and English dialogue in a very narrow domain of topics, that is, conversations for the conference registration. The words to be covered are 1500, and speaker-independent speech recognition technology is being developed. They are making steady progress, but we have to wait another ten years for a successful demonstration. Matsushita Co., NEC, Fujitsu, Toshiba, and some others are doing similar research.

SUMMARY

In summary, machine translation in Japan is in transition from research and development into commercial products and practical usage. Machine translation and natural language processing technology is regarded as a key technology in the twenty first century. It is recognized that the cooperation between developers and users of machine translation systems are essential for improved systems and wider customers. A Machine Translation Summit Meeting was organized in 1987 in Japan to offer a chance for developers, users, researchers and government officials who support R&D to exchange opinions and look for the future. It was very successful, and as a result the second and third MT Summit Meetings were held in München and Washington DC in 1989 and 1991 respectively. Through these three conferences we discussed the establishment of the International Association for Machine Translation. It was started in July 1990 at the MT Summit Meeting in Washington DC. It has three regional associations, namely European Association for Machine Translation, Association for Machine Translation in Americas, and Japan Association for Machine Translation. It will have various activities such as newsletter publication, conferences including Machine Translation Summit Meeting, workshops and seminars, tutorials, exchange of standard for dictionaries and text, and so forth. I hope, as the President of the International Association for Machine Translation, that people who are interested in machine translation join the association, and contribute to the progress of the systems and their use.

AUTHOR

Makato Nagao, Department of Electrical Engineering, Kyoto University, Japan