

Utilization of Machine Translation in Malaysia

Dr. Ahmad Zaki Abu Bakar
Malaysian National Institute of Translation
Kuala Lumpur, MALAYSIA

Abstract

This paper is a country report of the current progress of the research and development efforts in machine translation in Malaysia with emphasis on the work being carried out at the Malaysian National Institute of Translation.

1 Introduction

1.1 Malaysian MT Centers

Malaysia's involvement in machine translation started in the early 1980's with the establishment of its first machine translation center at the University of Science, Malaysia or USM in Penang. The center at USM was developed from collaborative research efforts with the University of Grenoble of France to build an English-Malay machine translation system. Later, USM moved to build computer-aided translation systems and the pursuance of other related machine translation studies through its own efforts [Zaharin'92].

The second Malaysian machine translation center emerged at the University of Technology, Malaysia or UTM in Kuala Lumpur. Based on a research and development laboratory working on natural language processing which started in 1981, the Computer Translation Unit later became the secretariat and main laboratory for the KANTA Project.

1.2 UTM MT Efforts

The KANTA Project is the Malaysian portion of the multilingual machine translation research project jointly carried out between Japan, Kingdom of Thailand, Republic of Indonesia, People's Republic of China and Malaysia. The project is better known as the CICC Machine Translation Project due to the Japanese Center of the International Cooperation in Computerization's role as the coordinator of the project.

Since 1987, Malaysia and the other four member countries have been attempting to develop an interlingua-based machine translation system. The first phase of the project officially ended at the end of March 1993. A prototype machine translation system which is capable of translating sentences in the national languages of the five member countries was able to be constructed.

Apart from the KANTA Project, the Computer Translation Unit at UTM is also pursuing other projects. A major ongoing project is the development of an English-Malay translation system through a joint research with the University of Manchester Institute of Science and Technology (UMIST). The project which started in 1990 is partially supported by the British Council and the Malaysian Research and Scientific Development Council.

UTM too has its own MT projects. Among the projects classified as national projects are a software to translate formal letters called LETRA [A.Zaki & Hapidah 91] and a software to translate Malay cooking recipes [A.Zaki 92].

2 Malaysian MT Plans

2.1 Kanta Project Phase Two

As announced by the Honourable Minister of Education of Malaysia at the Third Verification Test organised by CICC in Tokyo on 15 December 1992, the Malaysian Government has consented to an extension phase of two years for the cooperative machine translation project with Japan.

During this phase, the prototype machine translation system will be improved in all aspects to enable its performance to be better or at par with commercial machine translation systems. The extension will provide ample time for each member country to develop the prototype machine translation system further to reach its full potential as planned at the beginning of the project in 1987. The extra time will also provide the cooperating countries with opportunities to develop its human resources and explore potential applications for the finished product.

2.2 Malaysia's Translation Institute

As also announced by the Honourable Minister of Education of Malaysia at the Third Verification Test, the second phase of the KANTA Project is to be carried out in Malaysia at the Malaysian National Institute of Translation (MNIT).

MNIT is being setup in Kuala Lumpur under the auspices of the Ministry of Education, Malaysia. MNIT is to act as the sole official organization in Malaysia to manage and conduct activities pertaining to translation, interpretation and information exchange at the national and international level on behalf of the Malaysian Government. The final approval to establish MNIT was granted by the Malaysian Government on 28 April 1993.

MNIT will concentrate to provide efficient translation services to the Malaysian public. To undertake such tasks, MNIT needs to utilize human translators as well as machine translation systems in an integrated translation environment. In this regards, research and development plays important roles in the production of efficient tools and systems to assist the translation process.

To take advantage of the valuable resources and experience in MT research and development that is available locally, the Malaysian Government decided to make the KANTA Project the basis of MNIT's research and development department. MNIT will also have the capability to be connected by computer networks to other translation centers in the region as was jointly developed by the Open Systems Interconnection (OSI) research group in the KANTA Project.

Beside functioning as a service and research center, MNIT will also function as a training and transfer technology center for translation and interpretation. MNIT will assist to train more human translators, interpreters as well as machine translation developers to fulfil Malaysia's demands to expand its translation capability. The establishment of MNIT will assist Malaysia to solve the language barrier hampering the easy flow of information to the Malaysian national language.

3. MT Utilization

3.1. Accomplishments Study

A study was carried out in Malaysia to identify potential areas where MT systems can be utilized at MNIT. The Malaysian multilingual machine translation steering committee, chaired by the Ministry of Education also directed the study to cover the whole spectrum of the translation field rather than limiting the study to machine translation alone. The rationale behind this decision is not to fall into the old trap of having solutions find problems. A better approach is of course, to study the problem and then attempt to solve them efficiently.

3.2. Book Translation

Based on this wisdom, the study conducted was first aimed to identify immediate problems faced by Malaysian that require translation. From various sources, several parties in Malaysia on various occasions have highlighted the need to have more books in the Malaysian national language.

The Malaysian national language policy advocate the use of the Malay language as the official language for all official and administrative purposes and functions in the country. The Malay language was also given additional responsibility as the medium of instruction in primary, secondary and tertiary levels.

This policy as well as the national education policy require more books to be written in Malay. As a developing country having limited resources as well as due to the information explosion phenomena, the number of books produced by Malaysian authors are still insufficient to fulfil the demands of the Malaysian population.

This shortage necessitate the use of translation to extract knowledge and information from foreign languages. Dewan Bahasa dan Pustaka, the national agency setup by the Malaysian government to plan and develop the Malay language, estimates 1,000 to 5,000 books should be translated annually. Currently, only 78 academic books could be translated annually by the Dewan Bahasa dan Pustaka.

The foreign books to be translated have mostly been in English. However, due to the shift of knowledge centers of science and technology to non-English industrial countries like Japan, the English language monopoly has been broken. Malaysia now also translates books in French and Japanese. In fact many comic magazines as well as television series translated from Japanese are now in the Malaysian market.

3.3. Machine-Aided Book Translation

Studying the books to be translated, we found several problems in attempting to automate the translation of books. Serious problems identified are due to the abundant ambiguities as well as the varying styles and usage of language of different authors even on the same subject. Many concepts and terms in the new books are not available in the current Malay dictionaries.

Based on our knowledge of the state of the art of machine translation, we have to conclude that currently it is impossible to translate books automatically especially for an interlingua-based system. A great deal of pre-editing and post editing work will have to be carried out. The time consumed for these tasks will be more than the time used by human translators. The high cost of translation will also discourage the use of machine translation.

The translation qualities produced by human translators are high. However, the time taken to produce the translation is long and the cost is still expensive. Furthermore, the subject experts to translate the books are few in numbers and they just don't have the time to translate. To solve these problems, machine assisted human translation systems or translator's workstations have a higher chance of success. Here, the human translates while the computer assists by providing facilities such as on-line dictionaries and thesauri as well as word-to-word or phrase-to-phrase translations. Implementations of this idea has been carried out successfully by USM and MNIT will utilize such systems in its book translation activities.

3.4. News Bulletin Translation

Having found automatic machine translation systems to perform poorly at book translation, the study moved to find other potential areas. Our search found many opportunities in news bulletin of specific areas. A typical example is the stock exchange. Currently, a daily summary of the activity of the Kuala Lumpur Stock Exchange (KLSE) is published in the local news paper. Interests to have a more regular summary at specific intervals and in several languages are high. Foreign investors such as from Japan have shown their interest to have regular summaries of the KLSE activities in languages such as Japanese.

Careful study of the text corpus to be translated indicated that stock news bulletins are written in a sub-language. The concepts and terms used are almost stable. The frequent variables are normally the names of the new companies listed in the stock exchange board. The stock exchange activity summaries are also normally written in English and then translated into Malay and other languages. We found this situation to be an ideal application of the machine translation system we have produced through the CICC project. To further excite us, several news bulletin systems and value-added network systems are already available in Malaysia that will permit our translation system to be connected to their systems.

To experiment with this domain, we have been attempting to translate the KLSE daily summary first by hand and then by machine. At the moment we found the results of our experiments to be encouraging. We hope to improve the system in the second phase of the KANTA Project. If all of the member countries of the CICC multilingual machine translation project can agree to cooperate on the same tasks, then it would be possible to view summaries of the stock exchange in the national language of the cooperating countries.

4. Conclusion

Through our study, we have managed to find a *niche* for our machine translation system. The stock news bulletin is a good domain to experiment the capabilities of machine translation. In Malaysia, the research guidelines drawn by the Malaysian government necessitate that research and development projects must have immediate use to the country. As a developing country we have to forego the luxury of doing research for the sake of research. Here, we not only found an immediate use at MNIT for our six years labour but also an opportunity to educate the Malaysian economic sector that machine translation systems can have a significant place in this sector.

5. Reference

[A.Zaki 92] Ahmad Zaki Abu Bakar, "A cooking recipe natural language system", *Proc. International Symposium on Natural Language Understanding and AI as part of International Symposia on Information Sciences*, pp. 251-254, Iizuka, Japan, 12-15 July 1992.

[A.Zaki & Hapidah 91] Ahmad Zaki Abu Bakar and Noor Hapidah, "Malay official letters translation", *Proc. International Conference on Current Issues in Computational Linguistics*. pp. 413-420, Penang, Malaysia, 10-14 Jun 1991.

[Zaharin 92] Zaharin Yusoff, Vincent Gregory and Tan Yee Chum, "On the use of translators workstations and user-driven MT towards Meeting Today's Translation Needs in Malaysia", *Proc. International Symposium on Natural Language Understanding and AI as part of International Symposia on Information Sciences*, pp. 237- 244, Iizuka, Japan, 12-15 July 1992.