[From: Makoto Nagao (ed.) *Machine translation summit*. Tokyo: Ohmsha, 1989]

4 • 1 Commercial Machine Translation Systems

Hozumi Tanaka
Tokyo Institute of Technology

Since machine translation systems have passed through various research stages, several systems for commercial purposes have been put on the market. Although areas of application are limited, machine translation systems have been used in actual business situations for some time, and a variety of information concerning computerized translation has been accumulated. Such knowledge is undoubtedly valuable in developing better, easier-to-use systems. In this sense, it can be said that the research and development of machine translation has now entered a new phase.

This session was planned for the purpose of having developers of some of the representative commercial systems which have been used in actual situations talk about the concepts and technology adopted in developing the system as well as future research themes.

In parallel with this session, a demonstration session was held in the international conference hall displaying these machine translation systems. It is believed that the demonstration was useful in that conference participants were able to use these systems themselves and better understand the content of this session. Note, however, that of the systems summarized below, a demonstration of the METAL system was not given and that a demonstration of the DLT system was canceled due to the failure of the machine prepared by the conference organizer. Two systems developed by Oki and Mitsubishi, Pensé and Meltran respectively, were not explained here but demonstrated in the international conference hall.

ATLAS II, developed as a sequel commercial system to ATLAS I based on syntax processing, emphasizes semantic processing. As it is intended to be a multilingual translation system, structuring is done in such a way as to make analysis result rendition as universal as possible and not dependent on a particular target/source language. The importance of (a) identifying ways of effectively utilizing the existing machine translation systems, although it may require human editing before and after machine translation, (b) the need for enhancing not just information quantity in the dictionary but quality as well, (c) for integrating a world model, and (d) for further increasing computation capacity were pointed out.

HICATS/JE System handles the translation of text of chemical fields. Semantic processing is of particular importance. Analysis results are called dependency graphs, which are devised to be as independent from the target/source language as possible. The portion which depends on language passes through a transfer process. Moreover, an easy-to-describe format is adopted for the grammar description language. The importance of pre-editing, approaches to the sublanguage or control language, and the tools supporting them was pointed out.

A specific, detailed explanation of the basic concept of lexical transition network grammar introduced to develop large-scale grammar and semantic analysis rules for the development of the English-to-Japanese machine translation system TAURAS was given. It was maintained that the reduction of volume of calculation for analysis and generation was also realized. The system works on a translation workstation with a built-in English/Japanese editor system so that correction of translation results is facilitated.

PIVOT System adopts the interlingual method aiming at the realization of a multilingual machine translation system. The interim language used in the

PIVOT system was briefly explained. The interim language consists of basic units called CP. It also contains information about topics, focuses, and so on. In connection with problems of the present interim language, the importance of refinement of CP and pragmatic treatment of information was mentioned.

SYSTRAN, with a long history, presently develops machine translation systems for 15 language pairs covering a wide variety of fields from technical documents to literary works. The translation speed of the large model is as fast as 2 million words per hour. The system is structured in such a way as to effectively solve the problems of homonyms and ambiguity. Another characteristic of the system is relatively easy system expansion.

Siemens has been conducting the development of a German-English translation system METAL on a LISP machine, in collaboration with the University of Texas. Sentences to be inputted from an OCR are edited prior to machine translation; the post-editing of translation results are done on several PCs linked to the LISP machine. The translation software has a modular structure and adopts a transfer method. The importance of establishing solid machine translation theory and of semantic processing for future development was pointed out.

The DLT System adopts Esperanto as its IL (interim language), which enables it to conduct multilingual translation. Equipped with a knowledge base, the system solves ambiguity with the use of artificial intelligence; it is also capable of spotting ambiguity and requesting its elimination when a sentence is being input. Recently the DLT system has been improved by conducting the Melby test which examined the system's ability to solve word ambiguity. Accumulating technical terms into the IL for future use is now being considered.

To sum up, the following can be said about the above mentioned systems. Computers used for machine translation are either large model or workstation type. Neither type is capable of perfect translation at the moment. The systems limit the fields of translation and have post- and pre-editing functions in dialog mode. Due to the remarkable improvement of workstation functions in recent years, an increase of workstation-based systems is expected in the future. Many systems are aiming for multilingual translation, and the attempt to move closer to pivot language systems is noted. Attempts are also being made to positively incorporate artificial intelligence technology. The consistency of technical terms in translation results is frequently cited as one merit of machine translation. Key areas for future development are considered to be semantic and pragmatic processing, amount and type of data contained in the dictionary, and joint development of a large-scale dictionary. The view that securing a cooperative user who has system development ability is important for the development of a better machine translation system is noted.