

[*Machine Translation Summit*, September 17-19, 1987, Hakone Prince Hotel, Japan]

RESEARCH AND DEVELOPMENT PROJECT OF MACHINE TRANSLATION
SYSTEM WITH JAPAN'S NEIGHBORING COUNTRIES

YOSHIHIDE TSUJI
EXECUTIVE DIRECTOR
CICC

[**OUTLINE OF THE PROJECT**]

Research and development of machine translation system will be conducted with Japan's neighboring countries for interchange of information.

This project will undertake the research and development of machine translation systems, based on an optimal interlingua system. The specifications of the machine translation systems to be developed by this research cooperation project are as follows:

- 1) Translation fields :
Mainly industrial and technical information
- 2) Translation languages :
Machine translation between Japanese and Thai, between Japanese and Malaysian, between Japanese and Indonesian, and between Japanese and Chinese.
- 3) Translation accuracy :
80 to 90% (with pre-editing)
- 4) Term of the project :
6 years from 1987 to 1992
- 5) Elements of research and development :
Formulating of basic plan, Interlingua, text analysis system, Text generating system, Electronic dictionary system, Input/Output system, Translation support system, Integrated system, Standardization, Operation and dissemination
- 6) Evaluation and report of the result

1. Background to the Project

Japan has organized a cooperative research project named "Research Cooperation on Machine Translation Systems with Neighboring Countries" in 1987. We believe this project will bear great significance to the economies and societies of neighboring countries particularly in the following aspects:

- (1) Machine translation systems are aimed at lowering translation costs, cutting translation time, and dramatically improving translation capacity and they are expected to fill the gap between supply and demand. Japan's neighbors who are especially in need of technology transfer from Japan will be able to utilize machine translation systems for translation of industrial and technological data into their respective languages - contributing to greater technological and cultural exchange between Japan and these countries and to their economic growth.
- (2) In addition to the direct effects of machine translation systems, cooperative research is expected to provide assistance to the enhancement of research capabilities indispensable to neighboring countries for development - specifically speaking, in the form of service in an extensive area of basic technology, promotion of effective transfer of information processing technology and subsequent training of engineers expected to lead the growth in sophisticated information technology fields in their respective countries and to stimulate development of information processing technology in their native languages.

2. Objectives of the Project

- (1) To promote greater exchanges of technology and culture between Japan and its surrounding countries.
- (2) To stimulate effective transfer of information processing technology to Japan's partners in the joint R & D effort. To promote the effective functioning of information processing technology on the part of Japan's partners in joint research, and to cooperate with the training of technical personnel who will serve as the nucleus of the informationization process in the future.
- (3) To lend support in an extensive area of basic information technology through development of the technology in the respective native tongues in this R&D project.

3. The Goals of Research and Development

Four goals have been prescribed for this R & D project to encourage cultural and technological interchange in this area of the world and to assure greater development among the neighboring countries. The project is scheduled to be executed from 1987 to 1992 in the form of a cooperative effort with research institutes in neighboring countries (China, Thailand, Malaysia, and Indonesia).

- (1) The project will involve research and development of two-way machine translation systems between Japanese and Chinese, Japanese and Thai, Japanese and Malay, and Japanese and Bahasa Indonesian. These will be multi-interlingua translation systems employing an interlingua. Translation between the languages, such as between Chinese and Thai, is to be made possible at a level comparable to that between Japanese and other languages through addition and revision of grammar rules and dictionary definitions in the analysis and generation stages.
- (2) The area of translation covered by this project is to be documents concerned with industrial and technological data, such as technical bulletins and manuals. The number of basic terms to be used in the machine translation system is 50,000 and roughly 25,000 technical terms (in information processing). Translation accuracy to set at 80 to 90 percent with the use of technical terms dictionaries necessary (for literature concerning industry- and technology-related data in accordance with specific guidelines). The projected translation speed is more than 5,000 terms per hour.
- (3) The emphasis is laid on research on sentence analysis and generation, as well as dictionaries for basic terms. Development of technical terminology dictionaries, input/output translation support system, and integrated system will be carried out to improve operation efficiency and operability of the entire system.

4. The Elements in R&D

The elements to be studied are the following:

- 1) Interlingua
A multi-interlingua machine translation system employing an interlingua will be developed to make possible translation between any two languages other than Japanese as a potential capability, rather than focusing only on machine translation between Japanese and other languages.

- 2) Text analysis system
A text analysis system for morphemic, structural, and definition analyses on input sentence in a local language based on analytic grammar rules and conversion into an interlingua will be studied and developed. The study will also include an analytic grammar support system for support in writing and maintaining grammar rules.
- 3) Text generation system
Research and development will be conducted for a text generation system for style, structure, and morpheme generation on an interlingua based on generative grammar rules and conversion into the local language in question. It will also include study of a generative grammar support system for support in writing and maintaining grammar rules.
- 4) Electronic dictionary system
An electronic dictionary system for controlling grammar and definition data, as well as others, by recording equivalents between the local and interlingua will be under research. The number of terms are 50,000 basic terms and 25,000 terms in terminology dictionaries for information processing terms.
- 5) Input/output system
R&D will also be conducted on a keyboard input system and OCR (Optical Character Reader) for input and output of text and tables and graphics with text and an output system such as a printer for output in the local language.
- 6) Translation support system
Research will be done on support systems from a variety of approaches, such as pre-editing, post-editing, and interim-editing.
- 7) Integrated system
A machine translation system integrating the functions of the analysis, generation, electronic dictionary, input/output, and support systems will be developed.
- 8) Standardization
Standardization of the language codes in analytic and generative grammar and dictionaries and of the multi-language text base code system will be implemented.
- 9) Operation and popularization
To help with management, maintenance, and popularization of the machine translation system after completion, necessary software will be developed, and people trained.

Segment	1987	1988	1989	1990	1991	1992
1) Formulating Plan	Basic Plan		Fisical			
2) Interlingua	Trail Product	Development		Review		Evaluation
3) Text generation & analysis system	Basic design	Detail design		Development		Improvement
4) Electronic dictionary system	Basic design	Detail design		Development		Improvement
5) Input & output system	Basic design	Detail design		Development		Improvement
6) Translation Support System	Basic design	Detail design		Development		Improvement
7) Integrated system	Design	First-Product		Second-Product		System-Product