

SYSTRAN JAPANESE SYSTEMS

Eriko Akazawa
Systran Corporation, Tokyo, Japan

Japan has been criticized by the Western world that she is so anxious in exporting her technical goods while she seems rather reluctant in exporting information regarding her technology. Aside from the debate whether this criticism holds true or not, the fact is that more information is coming into Japan than going out of Japan.

We believe that one of the main reasons is simply that Japanese is very difficult to translate because it is strikingly different from the languages in the Western world. We are the only one in the developed free industrial nations whose native tongue is not a member of the Indo-European language family. For most of you who are here today and who live in the places where more than one language is used and understood, it may be interesting to note that Japanese is only used in Japan and no other places in the world use it as an official language (Fig. 1). Despite the uniqueness of the language, Japan's economy depends upon imports and exports with the rest of the world, which necessitates international communication. In addition, information regarding Japan, especially her technology, is very much wanted from the rest of the world nowadays.

This explains why the need for translation concerning the Japanese language is so high inside and outside of Japan and why the market for machine translation is very active in Japan now. Many Japanese mainframe computer makers, research institutes and translation companies are developing their machine translation systems and some are already on the market (Fig. 2).

It seems as though the Japanese finally saw machine translation as an answer to better and more timely international communications for overcoming the language barrier which surrounds the Japanese language.

Systran Corporation was founded in Tokyo in 1980. Mr. Kawasaki, president of the company, who is very sorry that he could not come to this conference, saw the value of the Systran system in overcoming the language barrier especially for world peace and undertook its extension to the Japanese language, under the supervision of Dr. Toma, entrusting development work to World Translation Center Inc. I was a member of the development team for both the English-Japanese and Japanese-English systems there. By 1984, both systems were brought close to an operational level and delivered to our corporation in Tokyo for further development and marketing.

Development of the Systran Japanese systems was not easy, because, as I mentioned earlier, the Japanese language is strikingly different from the Indo-European languages such as English which were previously covered by Systran. It is related neither to Chinese nor Arabic for which pilot Systran projects had been going before the Japanese project started in 1980.

I would like to take a few examples here to demonstrate some unique characteristics of Japanese which needed to be considered and taken care of in developing machine translation systems between Japanese and the other totally unrelated languages such as English (Fig. 3).

1. Japanese is written in a combination of Japanese alphabets, Kana, and Chinese characters.

As you can see, they do not look at all like the alphabets you are all familiar with. We need to have input/output devices to read, process and print these characters.

2. There are no spaces between words.

In order to analyze Japanese sentences, Systran must know where each word starts and ends in a sentence. The Word Boundary Analysis program was added to execute before the usual Systran translation process to do this job in the Japanese-English system. The WBA program analyzes input text utilizing a dictionary automatically made from the Systran Stem dictionary along with a series of rules to separate words. The accuracy of separation is now as high as 95%.

3. There are no special distinctions between singular and plural forms in nouns. No articles are added either.

For example, the Japanese word 'HON' can mean a book, books, the book and the books. Going from English into Japanese, this is not a problem. However, going in the opposite direction, the system must analyze whether 'HON' is plural or singular by looking for a modifier which implies plurality of the word in a sentence, for example, and generates the appropriate form.

4. Some verbs change form according to whether the object is animate or inanimate.

While there is no gender specification, the above specification is important in Japanese. This is a problem in the English-Japanese system, but not vice versa. For example, the verb 'TAKE' has to be checked for animateness of its object in order to be translated correctly.

This problem can be taken care of relatively easily by writing a rule in the contextual Systran dictionary.

5. Japanese classification of words is different from English.

English adjectives sometimes correspond to Japanese verb forms and so do English adverbs. Since there are no prepositions in Japanese, particles and verb forms often correspond to English prepositions. The Japanese equivalents to English prepositions change their forms according to what type of words govern them. Therefore, in translating a sentence, for example, I SAW A MAN WITH THE TELESCOPE, the English Analysis programs must set up the correct relationship between the preposition and its governor in order to have decent Japanese output.

6. Subjects and pronoun objects are often omitted.

This is quite a problem in the Japanese-English system. Since these components cannot be omitted in English, the Japanese Analysis programs must study syntactic and semantic information in detail to search for any clues on implied subjects and objects. One of the ways to deal with this difficult problem is to passivize Japanese active voice sentences without a subject in the English synthesis programs.

7. Japanese word order is almost a mirror image of English word order.

Since most of all words need to be rearranged, relationships among all words in a sentence must be set up as precisely as possible. This is because rearrangement rules are dependent upon the relationship set in the analysis programs. Therefore, success and failure of analysis of the source language show clearly in the word order of the target language.

This holds true for both Japanese-English and English-Japanese systems.

Although extra considerations were needed to be given in developing the Systran Japanese systems due to the language's idiosyncratic characters, we are happy to say that extension of Systran to the Japanese language has been very successful. We think that this is another demonstration of universality of the Systran systems.

In Japan, one of the largest translation centers in Tokyo has been supplying its customers with post-edited Systran English-Japanese translation output since the end of 1984. They had more than 10,000 English pages translated into Japanese in 1985 and expect to multiply the amount in 1986.

At the end of last year, we signed a licensing agreement with the U.S. Government for their use of the Systran Japanese-English system so as to expand their Japanese Information gathering capacity. We also hope very much to hear from the EC Commission's Japanese Information project soon of their start in utilization of the Systran Japanese-English system.

While the machine translation market is very active and we now have quite a number of competitors as I mentioned earlier, we believe that Systran has a leading edge over all of them.

The reasons are three-fold (Fig. 4).

First, the Speed.

When run under Facom M-380, the English-Japanese system translates up to 2 million words per CPU Hour and the Japanese-English system up to 1.5 million words per CPU Hour.

This speed means that Systran translates 25 times to 700 times faster than other Japanese MT systems. When you consider computer time used for machine translation and its related cost, the speed itself guarantees Systran's superiority in dealing with the large amount of texts from the economical viewpoint.

Secondly, Lenient Hardware Requirements.

While most of all the other MT systems are designed to work under specified hardware only, Systran can run under any IBM compatible machines such as Facom, Hitac and Siemens, with a minimum of 700k byte main storage area. Please note that IBM and IBM compatibles together have close to 90% of the world mainframe computer shares and 60% of the Japanese mainframe computer shares.

Last but not least, Sophisticated and yet Flexible Linguistic Logic.

In the Systran translation process, the analysis programs examine the text thoroughly from morphological, syntactic and semantic levels and reduce the information into an abstract Systran denominator which transcends the human languages. The synthesis program then converts the denominator into the specified target language. The framework of its logic is not only sophisticated but also flexible so that any rules for analysis and synthesis of languages can be incorporated into Systran to the extent that a linguist is able to formulate the rules explicitly and precisely. Furthermore, as mentioned earlier, universality of the Systran linguistic logic has been shown by its successful application to a number of different languages.

- 5 -

I have talked about the uniqueness of the Japanese language, the active translation market in Japan, development of the Systran Japanese systems and how they are utilized now. I have also explained what makes us believe in Systran's superiority over other systems. Now, I would like to conclude my presentation by talking a little about the future of Systran.

As an operational system, Systran must keep improving as feedback from users is incorporated into the programs and dictionaries. In addition, as Systran is an open-ended system, I believe that it will keep growing as it incorporates new knowledge and techniques in computer technology, linguistics, artificial intelligence, cognitive science et cetera. But above all, what seems to me the most important and indispensable thing for the future of Systran is collaboration among all Systran developers and users worldwide with close communication. We have gathered here today from different parts of the world to exchange opinions regarding Systran, our common treasure. Let us work together and keep Systran the number one machine translation system in the world.

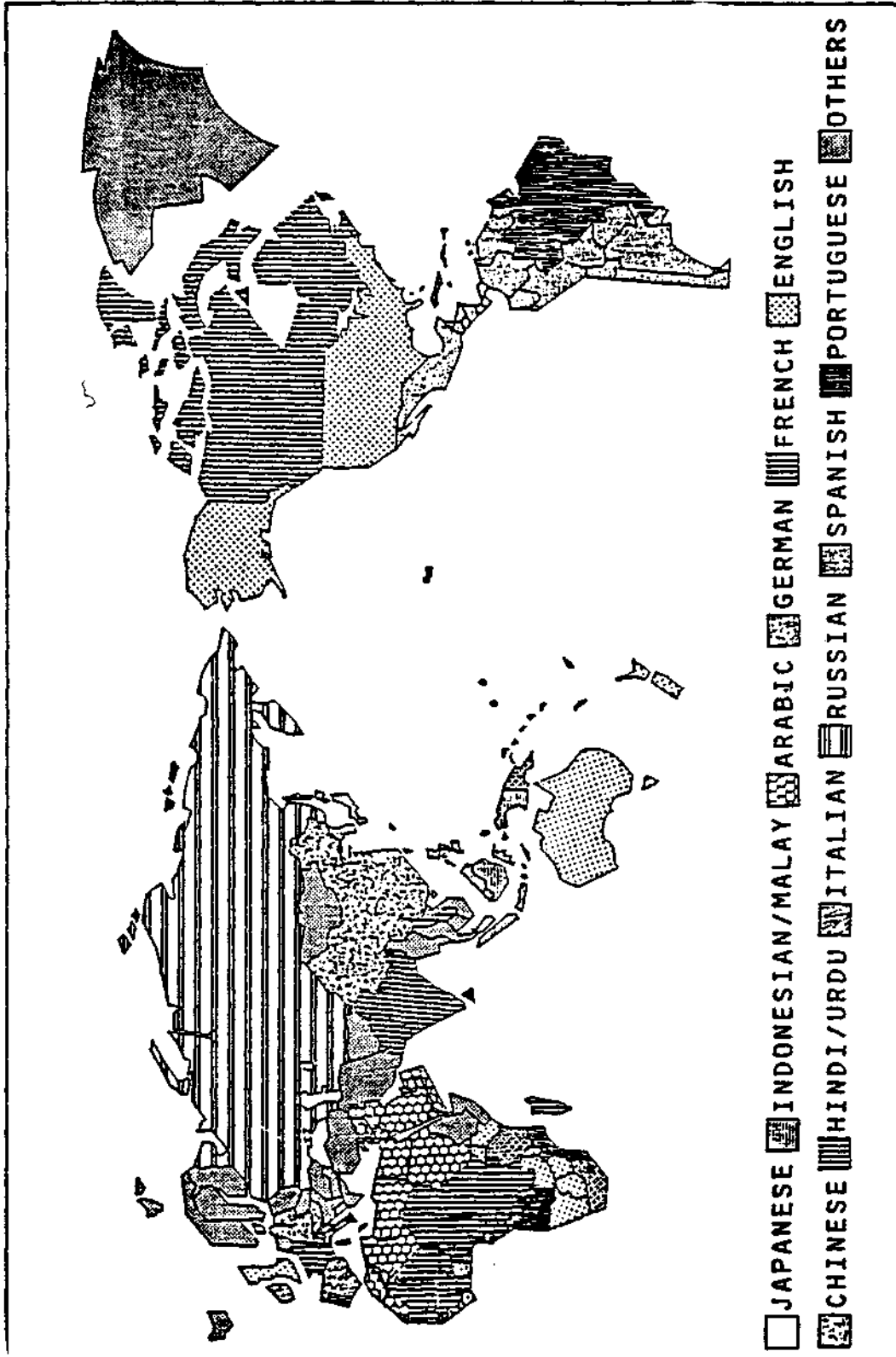


FIG.1 WORLD LANGUAGE MAP

< MACHINE TRANSLATION SYSTEM DEVELOPERS IN JAPAN >

SYSTRAN CORPORATION

FUJITSU LIMITED

HITACHI LIMITED

TOSHIBA CORPORATION

NEC CORPORATION

IBM JAPAN, LIMITED

BRAVICE INTERNATIONAL CORPORATION

NATIONAL MACHINE TRANSLATION PROJECT (JICST)

MUSASHINO ELECTRICAL COMMUNICATION

ELECTRO-TECHNICAL LABORATORY (ETL)

LABORATORY OF INTERNATIONAL TELEPHONES AND TELEGRAMS

ETC.

FIG.2

シストラン世界大会に出席するために、世界各地から
シストランに興味を持つ人々がルクセンブルグに集まった。

IN ORDER TO ATTEND THE WORLD SYSTRAN CONFERENCE,
PEOPLE WHO ARE INTERESTED IN SYSTRAN GATHERED IN
LUXEMBOURG FROM DIFFERENT PARTS OF THE WORLD.

FIG.3-1,ii

本
(HON)



A BOOK, BOOKS, THE BOOK, THE BOOKS

沢山の本
(TAKUSAN NO HON)



MANY BOOKS

FIG.3-iii

I WILL IAKE MY SISTER WITH ME.



連れていく。

(TSURETEIKU)

I WILL IAKE MY CAMERA WITH ME.



持っていく。

(MOTTEIKU)

FIG.3-iv

I SAW A MAN WITH THE TELESCOPE.



望遠鏡を持っている男の人を見た。

(BOENKYO WO MOTTEIRU OTOKO NO HITO WO MITA)

I SAW A MAN WITH THE TELESCOPE.



望遠鏡で男の人を見た。

(BOENKYO DE OTOKO NO HITO WO MITA)

FIG.3-v,v i

PEOPLE INTERESTED IN SYSTRAN GATHERED HERE TODAY FROM ALL OVER THE WORLD
1 2 3 4 5 6 7 8 9 10 11 12



世界中から今日ここにシストランに興味を持つ人々が集まった。

SEKAI JU KARA KYO KOKONI SISUTORAN NI KYOMI WO MOTU HITOBITO GA ATUMATTA
12 9 8 7 6 4 3 2 1 5

FIG.3-vii