System Demonstration

JICST Japanese-English Machine Translation System

Patricia O'Neill-Brown

Asia Pacific Technology Program
U.S. Department of Commerce

1. System builders and contacts

Builder: Japan Information Center for Science and Technology (JICST)

Contact: Patricia O'Neill-Brown
Ph. 202-482-6805
Email: poneillb@doc.gov

2. System category

Commercially sold.

3. System characteristics

3.1 Expertise required by the user

The user must 1) read Japanese, since menus are in Japanese and keyboard input must be done in Japanese; 2) set up the Japanese operating system and MTS on the machine the system runs on; and 3) operate in a Japanese language computing environment.

3.2 Domains covered

Science and technology. The machine translation system (MTS) was developed in the 1980s through a joint project between Kyoto University and the Science and Technology Agency (STA). The MTS is built to translate scientific and technical documents into English. The Japan Information Center for Science and Technology (JICST) of the STA has been developing and using the system. JICST is a public corporation which was established by the Japanese government in 1957 as Japan's key center for scientific and technical information services. One of JICST's missions is to promote Japanese science and technology outside of Japan; specifically Japanese scientific and technical databases. JICST has been using the MTS to produce English-language scientific and technical databases from Japanese-language databases.

3.3 Translation speed:
15, 000 words per hour.

3.4 Input formats accepted

1) Japanese text can be typed directly into the MTS system through the keyboard. Japanese is input through the *romaji -> kana -> kanji* conversion method through the keyboard.

2) Japanese text files can be imported into the MTS system. The MTS only accepts Japanese text files which are encoded in SJIS. SJIS is a Japanese encoding method. Japanese is not encoded in ASCII. The most commonly used encoding methods for Japanese are JIS (Japanese Industrial Standard); SJIS (Shift JIS); and EUC (Extended UNIX Code). If you are receiving Japanese text files via the Internet, they should be encoded in JIS since JIS is the most stable encoding method for file interchange over the Internet. Since the JICST MTS system accepts only files encoded in SJIS, files which are encoded in JIS or EUC can be converted to SJIS using a handy code conversion tool developed by Natsu Sakamura. The name of this JIS<->SJIS<->EUC conversion tool is JDCONV-DD. It can be downloaded from http://jasper.ora.com/lunde/j_tools.html on the World Wide Web.

If the document to be translated is in paper format, it must be scanned using a Japanese Optical Character Recognition (OCR) program into a text file format. Japanese OCR programs which operate on PCs and Macs typically save files in SJIS format.

3.5 Charts, tables and graphs

The system does not handle automatic processing of charts, graphs, and tables; therefore the information contained in them must be input into the system by hand.

4. Resources and System Internals

4.1 Size of lexicon

The system contains a main dictionary and a user dictionary. In the main dictionary, 590,000 words are in the Japanese dictionary: 570,000 nouns; 15,000 verbs; and 7,000 adjectives. 430,000 words are in the English dictionary: 410,000 nouns; 5,000 verbs; and 7,000 adjectives. A user dictionary is available for entering nouns, verbs, adjectives and adjectival verbs not found in the main dictionary. To classify these words in the dictionary, users must answer questions of Japanese and English grammar that the system requires. The user dictionary can be set so that the English definitions entered there are used by default, overriding the definitions in the main dictionary.

4.2 System type
Semantic transfer J/E.

4.3 Grammar
1500 Japanese rules; 500 transfer rules; 450 English rules.

5. Hardware and software

Macintosh Version System Requirements:

Hardware Requirements:

Power Macintosh is preferable

HD: 300+MB

RAM: 16+MB

CD-ROM Drive

Software Requirements:

System 7 or System 7.5

Japanese OS (Japanese Language Kit or KanjiTalk)

Word Processing Program or Database Program which can handle Japanese Characters (Japanese SimpleText, Nisus, Ichitaro, Microsoft Word, Excel J, etc.)

PC Version System Requirements:

Hardware Requirements:

IBM-Compatible PC

HD: 300+MB

RAM: 16+MB

CD-ROM Drive

Software Requirements:

Japanese Windows 3.1J

Word Processing Program or Database Program which can handle Japanese Characters (Ichitaro, Works for Windows 95, etc.)

Internet Service Version Requirements:

Japanese documents can be emailed to JICST in JIS (Japanese Industrial Standard) format for automatic processing and delivery of the raw output back to the customer's email address.

6. Functionality description

6.1 Price and Availability

PC and Mac Versions:

The CD-ROM version of the system will be put up for sale in the Fall of 1996 for about $500 from JICST.

Internet Service Translation Fees:

One yen per character of the English translation output (plus 3% Japanese consumption tax.)

Example:

This is a pen —>4 + 2+l+3=10 characters => 10 Yen

6.2 The Machine Translation Center for Japanese Science and Technology Literature at the U.S. Department of Commerce

On May 1, 1996, Commerce Secretary Michael Kantor and Japan's Minister of State for Science and Technology signed an agreement establishing the Machine Translation Center for Japanese Science and Technology Literature at the U.S. Department of Commerce. The Center was established through joint cooperation with JICST and the Office of Technology Policy at the U.S. Department of Commerce. The Center uses the JICST Japanese-English MTS System.

The JICST system was chosen because it is uniquely designed to handle scientific and technical documents. In addition, JICST and DOC's Japan Technology Program have been partnering in many different ways over the years in disseminating Japanese scientific and technical (S&T) information in the United States. The purpose of the Center is to demonstrate the usability of machine translation as a tool to help ease the language barrier U.S. researchers face in accessing Japanese scientific and technical (S&T) information. Use of the system is open to anyone from a U.S. organization, including those from universities, companies, non-profit organizations, trade associations, institutes, government agencies and individuals. Translations are not to be resold or otherwise disseminated without prior consent.

Response to the service has been very favorable. Most requests that come in are for patents to be translated and almost all of the requests come from companies. Several of the people whom have had documents translated using the system have said that their intent in using the system was to get a rough idea of what the documents were about and that the output from the system did a good job at providing this for them. One user remarked: "while the translations were far from perfect, they gave us an idea of what the article was about, which is all that we needed."

This filtering mechanism that use of the machine translation system provides is a boon for translation businesses. Many people from industry are "turned off to having documents hand-translated because of the high cost of Japanese-English translations, especially if the documents turn out to be irrelevant. Machine translation, however, can bring more businesses to professional translators. If people use machine translation to separate the 'wheat from the chaff,' they are more likely to have the 'wheat' sent out to be hand-translated.

The Machine Translation Center at Commerce demonstrates the growing awareness and importance which companies are placing on machine translation as a tool in their activities. The competition in the international marketplace demands that companies be able to quickly and competently operate in a multilingual environment. Machine translation is an important element in meeting that demand.