Adaptation of JICST's MT System for Workstation and PC's

Tatsuo Ashizaki

JICST (The Japan Information Center of Science and Technology)

1. Introduction

In order to prepare English databases more efficiently in response to the increasing demand for Japanese science and technology information in the US. and European countries, the Japan Information Center of Science and Technology (JICST) has developed a practical machine translation system. In house development activity from 1986 to 1991 was based on the results of the "Research on Fast Information Services between Japanese and English for Scientific and Engineering Literature" (Mu-Project, 1982-1986).

The JICST machine translation system (JMT) has been used in the production of JICST's English database since 1990. More recently, in 1994-95, downsizing was undertaken; at present, workstation (WS) and personal computer (PC) versions have been completed. The WS version is used with a network for English DB preparation and translation of large input sentences. Individuals will use the PC version for the preparation of English abstracts.

2. Outline of System

The WS and PC versions use the same dictionary and grammar rules but differ with respect to the user interface and the entry of new words into a user dictionary. Some beneficial features of the PC (Mac) version are explained.

2.1 PC version operating environment
  •hardware:
    Macintosh 680x0 or Power PC system
    memory (RAM): 8MB minimum (16MB recommended)
  • software:
    Kanji Talk 7.1

2.2 Editing Processing

• Reading
Besides translating input directly from the terminal, the system can also read Japanese text files prepared on the Mac or in DOS. It is thus possible to carry out translation processing on text files taken from electronic mail, OCR, etc.

• Saving
It is also possible to save the English output in a text file, without the additional information from the Japanese-to-English translation, for use with other software applications.

• Printing
Japanese and English text may be printed out either together or separately.

• Sentence recognition
Text read during sentence recognition processing is converted from half width to full width characters. Processing is also carried out for itemization sentences, except for the (.) period, when they are separated into one sentence in order to translate the next.

• Editing
Editing functions are as follows:
cut, copy, paste, retrieval/change, combination of line, division of line, insertion of line, deletion of line, paste of recognized sentence, copy of sentence combination, display of clip board

2.3 Translation processing

It is possible to translate 200 consecutive sentences of up to 150 characters each.

• Whole sentence translation
Translation of all the sentences is initiated by a single command. As each sentence is translated, the result is displayed.

• Sentence number translation

Translation and execution are carried out in accordance with a selected translation processing range.

• One sentence translation
Only the designated sentence in a translation will be translated.

2.4 Corresponding relation

• Display of the translated word
Expressions in a translated English sentence that correspond to specific expressions in the Japanese input are indicated on the screen. Japanese is also shown in inverse video for a designated English expression. When more than two words are included in a designated character string in the translation, the head word is highlighted. If there is no word corresponding to a designated character string, the display is not highlighted.

• Change of the translated word
When alternatives are available in the master dictionary, the user can select a substitute in the translation result. Alternatives from the user dictionary, however, are not displayed.

2.5 System and User dictionary

• User dictionary
The user may specify a dictionary to be used for words not found in the (master) translation dictionary. Such a dictionary is called the user dictionary. If a preference for the user dictionary is indicated, words entered there will be displayed with the translation result. Nouns, verbs, adjectives and adjectival verbs can be added to the user dictionary.

• User dictionary operation
The word the user wants to add to the user dictionary is retrieved. A new entry is then made in the user dictionary, correcting the corresponding entry in the master dictionary. Thus, even though contents of the master dictionary itself are not changed, the user designated translation is displayed because the user-dictionary is given precedence in translation and processing.

• Previous user dictionary entries

Since only one user dictionary is permitted, new entries may be made to correct previously entered data. When an entry is not found in either the master or user dictionaries, it is added to the user dictionary.

2.6 Discriminative translation condition

The user can change the output sentence style. For example, if there is no subject in an input sentence, the first person pronoun ("I") may be selected as the subject for the English output. Other alternatives include designating "It" as the subject, or translating into passive voice or an imperative sentence.

3. Dictionary of Production/Management System

Using the UNIX-ADABAS database management system, a multi-window, GUI-type translation dictionary is prepared. As of May, 1995, the sizes of the machine translation dictionaries are as follows:

Japanese dictionary:	580,000 words
Nouns:	560,000 words
Verbs:	15,000 words
Adjectives:	7,000 words
English dictionary:	430,000 words
Nouns:	410,000 words
Verbs:	5,000 words
Adjectives:	7,000 words

4. Future Plan

Scheduled for completion in summer of 1995, a Windows version will have an operating method essentially identical to the Mac version. Network service, via Internet, will begin in the fall of 1995. For now, JMT will be used only within JICST for abstracting and translating, as there is no immediate plan to make the software itself available to the general public. I hope to provide this system to the public in the future.