[*International Conference on the State of Machine Translation in America, Asia and Europe*. Proceedings of IAI-MT86, 20-22 August 1986, Bürgerhaus, Dudweiler]

Unified Operation Control System for Mu Machine Translation

Yoshiyuki Sakamoto
Electrotechnical Laboratory
Sakura-mura
Niihari-gun
Ibaraki, Japan

Taeko Ariga Atsushi Washio
Facom-Hitac Ltd.
Tsukuba branch
Sakura-mura
Niihari-gun
Ibaraki, Japan

0. Introduction

Since 1982, under a state-subsidized project for developing a machine translation system called the Mu* project, efforts have been made to prepare the necessary basic and detail designs, software and data.

The project is intended to develop a unified machine translation system including a language processing system and a dictionary system for practical use. The unified system act as an interface between the language processing system, which is the core of the translation system, and users. In translating an original text, it can display the current status of translation and permits correction of the original text and its translation easily. It is also capable of translating a large volume of text collectively. Consideration is given to modifying the dictionary freely so that its inadequacies, if any, can be supplemented.

The unified Japanese-English machine translation system is scheduled to be completed in March. 1986, featuring easy use, versatility and expandability. It will be run on a large computer at the Research Information Processing System (RIPS) for practical use. This paper will discuss its configuration, environments and functions, together with the results of translation experiments.

In developing the system, it has been assumed that users of the system have no knowledge about the language processing system, that is, the translation software. Consideration has therefore been given to developing a system that can be operated by users easily and efficiently.

1. Purpose

The increasing sophistication of science and technology has been accelerating the accumulation of information and know-how. For researchers to do creative work efficiently, it is absolutely necessary for them to be able to gain access to the vast volume of state-of-the-art literature in broad fields of science and technology.

This calls for information services which provide flash reports on the scientific literature written in Japanese and in English. As one of these, there has been demand for scientific paper and abstract translation services as a support for RIPS users undertaking there research work. To meet the demand, joint efforts have been made by Kyoto University, Electrotechnical Laboratory (ETL), The Japan Information Center of Science and Technology (JICST), and Research Information Processing System (RIPS) since 1982 to develop a machine translation system (A Research on Japanese and English Scientific Literature System). There are already a number of machine translation system developed or currently under development both at home and abroad. The joint project referred to above is intended to develop an efficient scientific literature translation system to promote scientific exchange with foreign countries. Specifically, the following efforts have been made:

(1) Development of a language processing system to translate scientific abstracts written in Japanese and in English utilizing a scientific terminology data base (by Kyoto University and ETL)

(2) Development of a data base of Japanese and English scientific terms (by JICST and ETL)

(3) Development of a translation system incorporating the language processing system and data base above (by RIPS)

The following are the purpose of the translation system. There are two typical uses of any machine translation system;

(1) Collective translation of a large volume of text

To translate a large volume of abstracts, etc. on a batch processing basis for later editing by specialists

(2) Interactive real-time translation

To translate the text of a paper prepared using a Japanese or English editor or the like in an interactive manner to allow the author to edit both the original text and its translation on the CRT screen.

To achieve the two types of uses, the unified machine translation system has the following functions:

(1) Interactive editing of original text and its translation

Permits editing both Japanese and English sentences on the same display terminal.

(2) Interactive request for translation

Specifying the sentence to be translated on the

* This project is being carried out with the aid of a special grant for the promotion of science and technology from the Science and Technology Agency of the Japanese Government.

terminal screen enables the sentence to be translated and displayed on the same screen.

(3) Request for batch translation

Makes it possible to perform translation on a batch basis utilizing a large file.

(4) Interface with general-purpose Japanese and English editors

Besides, the following functions are provided for easy use:

(5) Interface with literature and information retrieval system

Provides interface with the existing literature retrieval system to permit translating titles, abstracts, etc.

(6) Editing and maintenance of dictionaries

Make it capable of editing the Japanese dictionary, English dictionary and Japanese-English/English-Japanese transfer dictionary on the display terminal

(7) Selection of dictionary based on the text to translate

Allows the user to specify private-use, technical terms, and/or basic dictionaries for appropriate translation.

(8) Maintenance of translation texts

Maintains original and translation text files.

(9) Image retrieval of dictionaries

Make it capable of creating a data base of dictionaries in the form of images to allow users to retrieve them as tools for translation.

2. Method of translation employed for the language processing system

The language processing system, the core of the Japanese-English (English-Japanese) translation system, has been developed in several parts, that is, Japanese (English) morphological analysis. Japanese (English) syntax analysis, Japanese-English (English-Japanese) transfer, English (Japanese) syntax generation, and English (Japanese) morphological generation. The morphological analysis and generation parts have been developed by ETL using UTILISP (University of Tokyo Interactive Lisp Processor). The syntax analysis, transfer and generation parts have been developed by Kyoto University using GRADE (GRAmmer DEscriber), a transfer syntax description language developed for this project.

For Japanese-English translation, 3.000 sentences picked up from JICST's files of scientific abstracts have been analyzed to work out analysis, transfer and generation rules and make a Japanese dictionary, and Japanese-English transfer dictionary, and English dictionary. As for English-Japanese transfer. As many as 3,000 sentences have been analyzed to define the corresponding rules and make the necessary dictionaries.

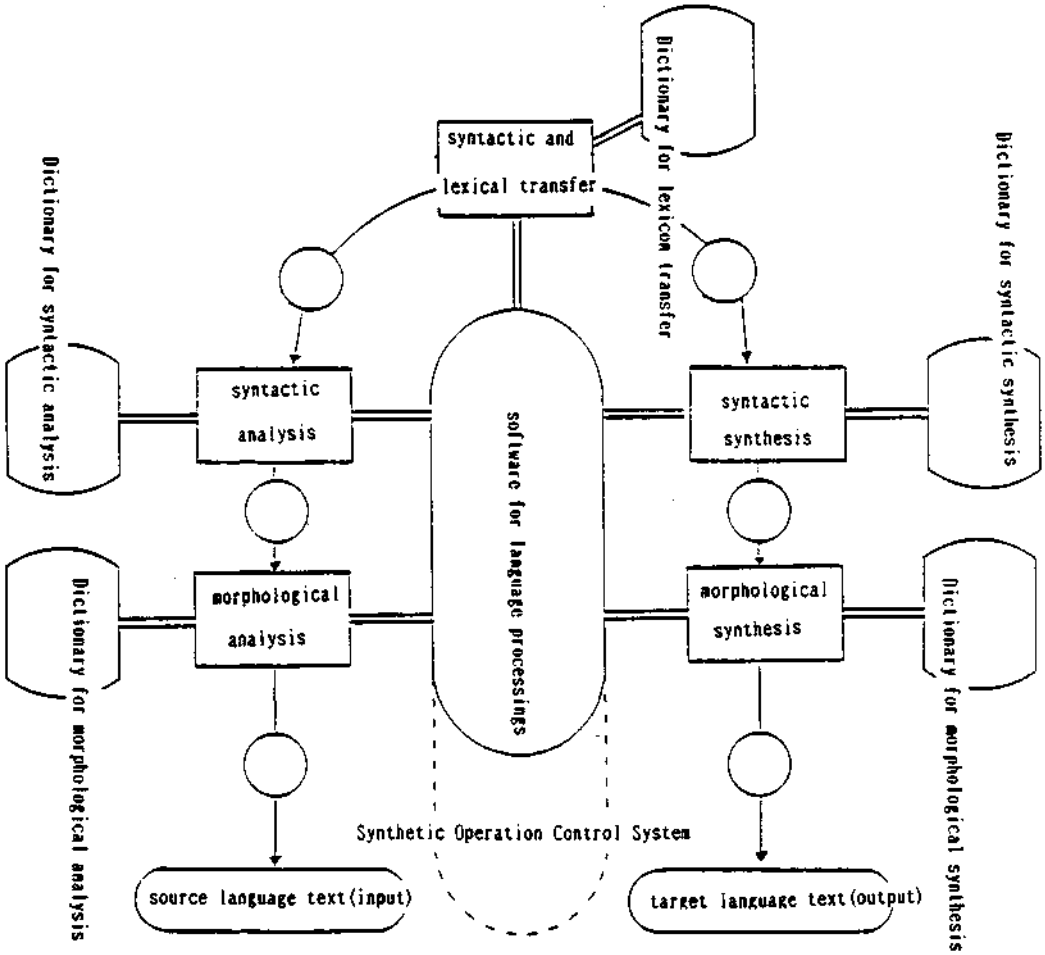

Fig. 1 System Organization for Mu Japanese-English Machine Translation

Characteristics of the language processing system

The system performs translation in five stages of processing, including morphological analysis, syntax analysis, syntax transfer, syntax generation, and morphological generation as shown in Figure 1.

As a Japanese-English translation system, this system has the following characteristics:

(1) This system does not require any human intervention in the stage of preprocessing. It makes a high quality translation direct from the original text. Use of the case grammar, which places emphasis on meaning, and capability of writing exception rules unique to produce a good translation.

(Note) The Japanese language, unlike English, French, German, etc., does not have strict word order. To analyze the syntax of Japanese sentences, therefore, this system employs not the syntactic grammar that places emphasis on word order but the case grammar that analyze the syntax of a sentence based on the cases of verbs and other words (e.g., agent, object, means).

(2) This system employs the so-called transfer method, with which a translation is made through the processes of analysis, transfer, and generation. And each of these processes employ the "partial grammar network" method. This makes it possible to apply the system to electric engineering and other specialized fields. Since the system is expandable and versatile, its application will not be limited to the area of scientific information retrieval.

A machine translation system in general translates the source language (Japanese in the case of Japanese-English translation) into the object language (English) using syntactic rules and a dictionary data base. Our system employs the transfer method for transfer processing.

The partial grammar network method is to classify a vast number of syntax rules into a number of interrelated groups. permitting specifying the scope of application of each rule in its group. With this, if it becomes necessary to add a new syntactic rule, it is possible to insert the rule in the appropriate place in the flow of overall processing while minimizing the influence to other parts.

(3) The Japanese-English scientific term data base used in this system is a highly reliable dictionary data base. It has undergone specialists' word-by-word checking.

It is expected that the Japanese, English and transfer dictionaries each will have 100.000 words when the project has been completed. Japanese dictionary entries include the nouns appearing in 3,000 sentences picked up from JICST's files of abstracts, and English in 3.000 sentences from INSPEC's files of abstracts. Japanese verbs are picked up from 13.000 sentences in JICST's files, and English verbs from 3,000 sentences in INSPEC's files. Besides, additional entries include about 50.000 nouns collected from the Japanese Scientific Terms and other sources.

This is one of the most reliable dictionaries available in Japan and can therefore be used not only for machine translation purposes but as a more general dictionary data base for standardization of scientific and technological terms, etc.

(4) Evaluation of the quality of translation

Method of evaluating the quality of translation

(i) Easy to understand:

Sample English translations output by the system have been evaluated in five grades by average native speakers to determine to what extent they can understand the translations without making reference to the original texts (Japanese).

(ii) Fidelity;

Sample translations have been evaluated in 7 grades by human translation experts to determine to what extent the output sentences reflect the information of the input sentences and to what extent the former are apart from the latter.

The system employs the concept of dictionary-based processing to handle the unique aspects of each language. With this system, it is easy to describe language information.

Sample Japanese-English machine translations

Samplel

この方法は，Ｒ．　Ｂｅｌｌｍａｎの理論から導いたもので，
解析幾何理論の応用である．

The method is one of derived from R.Bellman's theories and it is an application of analysis geometry theories.

Sample2

計算機で基板のパターン設計を行い，部品の手配から製品
までの各工程を制御するシステムを検討．

The systems are studied which carry out the pattern design of substractes by the computers and control each process from arrangements of parts up to products.

Sample3

実験と比較し，よい一致を見た．

The comparison was made with experiments, and good coincidence was obtained.

Sample4

緩やかな傾斜があり．凹凸のある表面によるラジオ波の熱放射．

The thermal radiation of radio waves by rugged surface with slow tilts.

In Sample 1, the input sentence was analyzed to determine whether the underscored 「で」 is a case particle or an auxiliary verb. The analysis found that the latter was the case so the input sentence was translated as two sentences connected by "and".
In Sample 2, the input sentence was analyzed to determine whether the verb 「設計を行い」 governs the noun 「システム」" or the verb 検討 (interpreted as the verb 検討する)" and it was decided that the former was the case. Since the input sentence has no subject, translating it required using the word "system" as the subject. Sample 3 is a sentence without a subject and has a difficult expression 「よい一致を見た」. Difficulty lies in translation 「一致」, a verb-derived noun, and the verb 「見た」 into "coincidence" and "obtain", respectively. The input sentence of Sample 4 is very difficult to analyze. The translation "with slow tilts" cannot be obtained unless it is decided that "... であり." do not separate the sentence into two but modifies 「表面

Specification of the language processing system

(1) Dictionaries

Japanese-English:
* Japanese morphological analysis dictionary
* Japanese syntax analysis dictionary
* Japanese-English transfer dictionary
* English syntax generation dictionary
* English morphological generation dictionary

Entries:
* Declinable words (verb, adjective etc.) ..... about 6.000
* Indeclinable words(noun, adverb etc.)..... about 70,000
* Auxiliary words 400

(2) Grammatical rules
* Japanese language analysis rules 1.700
* Japanese language transfer rules 900
* English language generation rules ... 630

(3) Language processing software

(i) Japanese-English translation software
* Japanese language morphological analysis program : 3.000 lines
* GRADE (syntax analysis, Japanese-English transfer, English syntax generation) : 15.000 lines
* English language morphological generation program: 1,000 lines

(ii) English-Japanese translation software
* English language morphological analysis program : 2.000 lines
* GRADE (syntax analysis, English-Japanese transfer, Japanese syntax generation) : 15.000 lines
* Japanese language morphological generation program: 800 lines

3.Configuration of the unified s ystem

The unified system can be divided broadly into a text editor, a dictionary editor, input/output text conversion, and system control and management. These correspond to the functions necessary to attain the purpose of the unified system. Figure 2 shows how these parts are related to the unified system. The system uses PL/I as its programming language for effective interface with the language processing system, for effective editing text and dictionary and for program maintenance. This results in efficient system operations and high-speed processing.
Each functions of this system has as following:

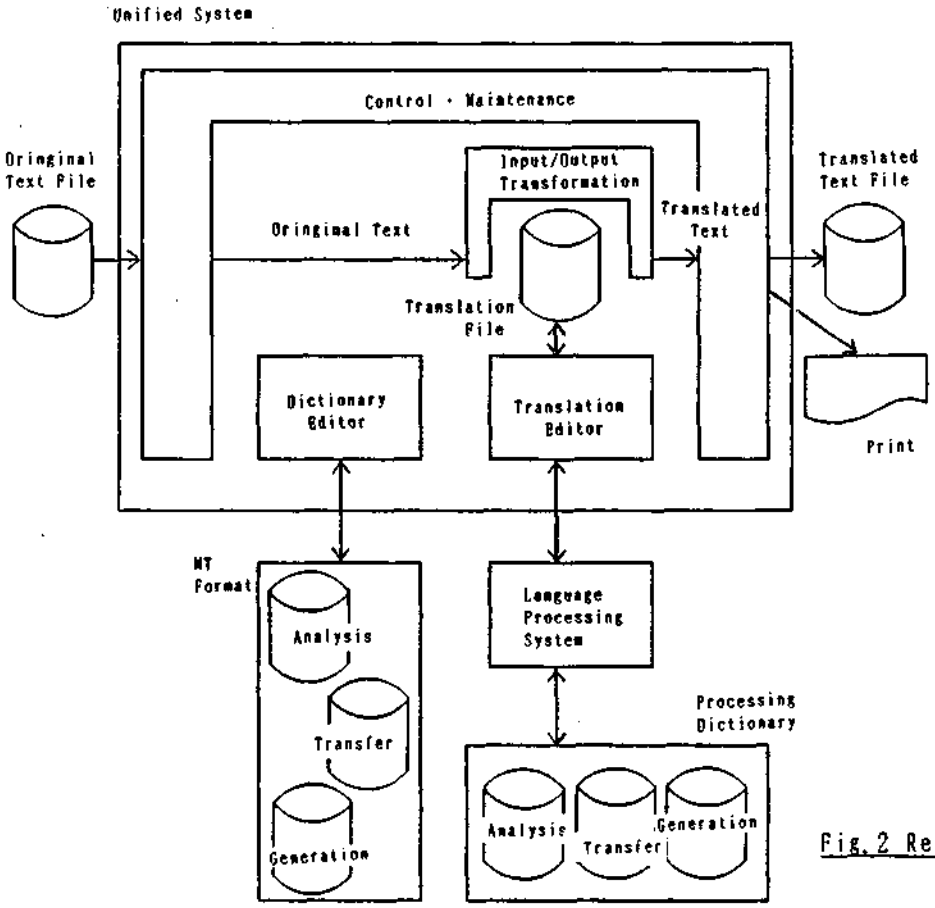

Fig.2 Relation of Parts in Unified System

(1) Text editor
Displays the original text and its translation on the CRT screen to edit then as needed. It is possible to start real-time translation processing. Interactive translation and editing operations are all performed on the text editor.

(2) Dictionary editor
Makes reference to and edits the contents of the dictionary.

(3) Input/output text conversion
Provides interface between the external Japanese and English editors and the literature retrieval system. Makes it possible input and output files of the formats unique to them. In input conversion, takes sentences one by one from the original text file and stores them into the translation file. In output conversion, writes the translated sentences in the translation file into the specific file.

(4) System Control and Maintenance
Provides control between the language processing system and the unified system, interactive control, batch translation requests, and file management. All the functions provided by the unified system are executed through conversation between the user and the control/management part.

(5) Translation internal file
The translation file is the file for translation and editing. It stores pairs of original text sentences and their translations by sentence number. This provides interface with the language processing system. Editing with the text editor is operated on this translation file. Sentence numbers are assigned from the first sentences of the original text file when its sentences are extracted for input conversion.
Original text statements and their translations are stored as pairs in one file. This makes it easy to relate original sentences to their translated sentences for editing purposes. Furthermore, when stopping the post-editing operation, it is unnecessary for the user to be concerned about the correspondence between the original text file and the file that stores the translation of the text. All the character data

in the translation file is expressed in the JEF code (kanji code to express one character in two bytes).

4.Translation processing functions

Normally, machine translation involves the steps of input of the original text, pre-translation editing, translation, post-translation editing, and output of the translated text. The unified system provides environments in which each of these processing steps can be performed efficiently.

4.1 Text editing

(1) Display of original and translated texts

The text is displayed sentence by sentence with sentence numbers. The original text is displayed on the upper part of the screen and the translated text on the lower part. The correspondence between the original text and its translation is given by sentence numbers. (Fig.3) The sentence means a single sentence taken out automatically by input conversion. An original sentence cannot be translated correctly unless the machine recognizes its meaning properly in advance. The result of automatic extraction of a sentence can be checked and corrected, if necessary, on the text editor screen.

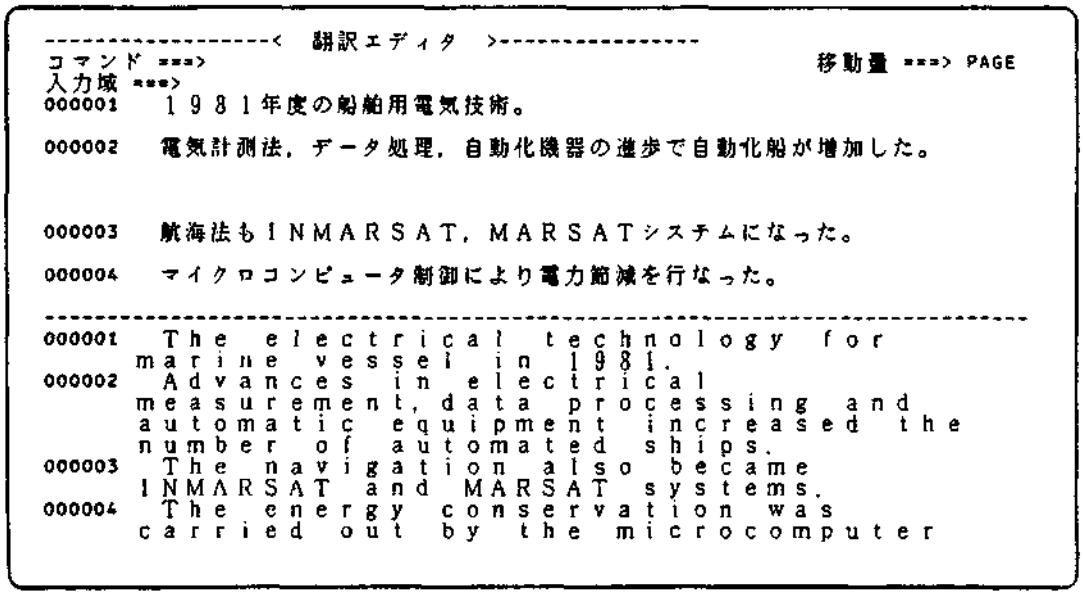

Fig.3 Display of the Translation Editor

(a) Sentence number

Sentence numbers are assigned internally for translation processing purposes and are independent of externally assigned information such as document numbers. These are important for translation, editing, output and printing. The text editor can divide, connect, delete and insert sentences. In editing that involves changing sentence numbers, it is necessary to maintain the correspondence between original sentences and translated sentences by means of sentence numbers. To this end, if a sentence number is changed in either the original text display area or the translated text display area, the corresponding sentence number in the other area is automatically changed accordingly.

If the translation file has no source text and it is necessary to input an original text from a terminal, sentence numbers are assigned sequentially starting from 1.

(b) Moving of displayed data (scroll)

Using the scroll key, it is possible to move the displayed text upward or downward by the specified distance. It is also possible to move to the beginning or end of a sentence or to the sentence number specified by an operand by means of commands. In these operations, both original and translated texts move in concert, so the display always shows the corresponding pair of original text and its translation.

(2) Editing

A character string in the display area can be modified directly. Character data are kanji codes. The keyboard does not have as many keys as all kanji codes. (There are more than 4000 kanji codes.) Characters are typed in alphabetical letters or kana and then converted into the corresponding kanji codes by means of the function key. The editor provides the following functions:

(a) Replacement of a character

(b) Insertion of a character

(c) Deletion of a character

(d) Input of a text

To insert a new text in the displayed text, specify the command in the sentence number in which the new text is to be included. Then a blank sentence is created, enter the text there. If the translation file has no sentence, a blank screen is displayed. In this case, it is possible to make the text sentence by sentence freely.

(e) Moving, copying, deletion of a text

By using the commands, it is possible to move and copy a text to a desired place. A command to delete a sentence is also provided.

(f) Separation and connection of sentences

It is possible to divide one sentence into two or connect one sentence with the subsequent sentence to make them a single sentence. This function is used when modifying the result of automatic sentence taking out, translation a long compound sentence in two sentences, or inserting a long character string in a sentence.

(g) Retrieval and change of character string

Retrieve the specified character string from the text and move the display screen to that place. Change the specified character string to another one.

(i) Printing of a text

Print the specified original text and its translation.

4.2,Input/output text conversion

Texts requiring translation may include texts prepared by an external editor, scientific papers, manuals and other machine readable documents. It is convenient if these documents can be input to the unified system directly, but it is impossible to provide the system with all file interfaces. The system is therefore provided with file interfaces that match with its operating environment.

(1) Input text conversion

This processing takes out one sentence from the original text file and inserts it in the appropriate place of the translation file in the form requested by the language processing system. Output files of external editors, and information retrieval systems and JICST abstract text files can be input as original text files. Other machine readable documents can therefore be translated if converted into these file formats.

(a) Extraction of a sentence

Extracting one sentence from the original text is the first step of the translation processing. A correct translation cannot be obtained if a single sentence is recognized wrongly. In the conversion of an input text, the end of one sentence is identified by a sentence separator. Sentence separators are a period and a question mark in the case of English and a period in the case of Japanese as default assignment. It may be necessary to use other symbols as sentence separators. This system allows any symbol to be specified as a sentence separator from the screen.

(b) Decision whether to translate a text or not

This system is primarily intended to translate scientific papers. Such document contains a title, an abstract of the content, and other bibliographical information. Translating all the information contained in the document does not necessarily produce a meaningful result, giving additional burden to the language processing system. In input conversion, therefore, a decision is made whether to translate the data of each abstract or not and the information showing to that effect (translation flag) is attached to each sentence. If a translation command is input, the text editor gives only the sentences with the flag 'T' on to the language processing system. For JICST abstract data, the translation flag is turned on for titles and excerpt texts. In output files of information retrieval systems, information varies with files. So, users specify the subjects. In input conversion, the translation flag is turned on only for the specified subjects. It is, however, desired to include bibliographic information in the translated text, even if it is judged as information not to be translated (with the translation flag 'NIL' set off) are written also in the translated text area of the translation file to ensure that the same information as the original text is conveyed.

(c} Formating control characters

A text prepared by an external editor includes control characters for formatting. A meaningful result cannot be obtained by translating a sentence that contains a control character which is meaningful only in the external editor. The input conversion of this system therefore excludes any control characters. External editor control characters may be changed. Giving consideration to this possibility, an external table of control character is provided while excluding any control character from the table.

(2) Output text conversion

Translated sentences are stored in the translation file. The output conversion of this system takes out only the translated sentences from the translation file and outputs them to an external file, serving as an exit file interface. Files that can be output are those handled by external editors and English runoff systems. It is also possible to print a translation and its original text side by side.

(3) Character code conversion

An original text to be given the language processing system and its translation output are both expressed in kanji codes (JEF codes). (An English sentence is, too, expressed in alphabet of the JEF code.) If an original text is in Japanese, no conversion is required. If, however, an original text is in English, it is necessary to convert the EBCDIC code into the JEF code. To output an English translation to a file for the English editor also requires JEF-EBCDIC code conversion. These conversions are performed automatically from the input/output file formats.

4.3 Interactive translation

Interactive translation is another major function of the translation editor. Besides, the text editor is capable of displaying the traces of translation. This can be used to improve the dictionary and check the process of translation.

(1) Interactive translation

Specifying the translation command to the sentence number of the sentence to translate causes the sentence to be transferred to the language processing system. Upon its translation, the translation result sentence is displayed in the place indicated by the corresponding sentence number in the translation display area. By specifying operands, it is also possible to display a termination message of each step and the current status of the processing.

The sentence to be translated are given to the language processing system as individual sentences identified by their respective sentence numbers. A correct result cannot be obtained if the same sentence number contains two or more sentences. It is necessary to be careful because execution of the sentence division command and/or the sentence connection command may affect the original sentence.

(2) Listing of unknown words

An original sentence cannot be translated properly if it contains unknown words. Translation of an original sentence will be facilitated if the unknown words contained in the sentence are identified in advance. To this end, the text editor is provided with the function of displaying unknown words, at the stages; (i) morphological analysis, (ii) generation of the specified. Unknown words after morphological analysis are mortal wound for translation. On the other hand, to get unknown words after generation is useful to know which the dictionaries have correspondence each other or not.

(3) Calling of the dictionary editor

When translation an original text with the text editor, it is often necessary to refer to or modify the dictionary. For example, adding the unknown words identified in the search of unknown words to the dictionary and modifying the definitions of dictionary entries for personal use. The dictionary editor can be invoked from the initial menu after terminating the text editor. More conveniently, it is possible to call the dictionary editor from the translation editor directly.

4.4 Batch translation

The batch translation function is used to translate a large volume of text on a file-by-file basis. The original sentences stored in the translation file after input conversion are translated on a batch basis. The results of the translation are output to the translation file, making it possible to check and correct the translated sentences with the text editor. A translated text can be printed and output to an external file with the output text conversion function.

5. Dictionary editing

A dictionary is one of the major factors affecting the quality of translation. For general users who cannot modify the core of the language processing system, the quality of translation can be improved only through modification of the dictionaries used. The basic dictionaries provided by the system give consideration to a maximum generality. But the contents of the dictionaries do not satisfy all users and the translations of terms may be different in some fields.

The unified system therefore refers to more than one dictionary for translation and provides a dictionary editor to allow each user to modify the contents of the dictionaries as needs.

5.1 Dictionary data base system (Fig.4)

(1) Processing dictionaries and MT format

In making a translation, the language processing system uses lexical rules for each word as local syntax rules as well as ordinary lexical information such as the translation of words, grammatical functions and meanings. Dictionaries of this type are called processing dictionaries. Different dictionaries are used in different processes of translation. including a morphological analysis dictionary, syntactic analysis dictionary, transfer dictionary, generation dictionary. These dictionaries are written in internal expression for ease of reference by the language procession system. It is difficult to edit them directly.

Dictionary information is accumulated in the form of the S expression of LISP called the MT format. The information includes the translations, grammatical functions and meanings of entry words. The processing dictionaries referred to above are created by the dictionary creation functions using the MT format. The dictionary editor edits the contents of a dictionary on the MT format. The results of the editing is reflected in the processing dictionaries created through the use of the dictionary creation functions. There are three types MT format, namely, analysis, transfer, and generation. From these, the five processing dictionaries are generated.

(2) Basic dictionaries and personal dictionaries

The basic dictionaries are general-purpose dictionaries provided by the system. These dictionaries contain basic words and technological terms mainly used by the in the field of electric engineering. A dictionary used by the user is called a personal dictionary.

Dictionaries should be steadily improved to meet the user's own requirements and to improve the quality of translation. It is, however, necessary to modify dictionaries. The well-balanced basic dictionary intended for general use may be distorted if user modifies it according to his/her own needs.

The basic dictionary may therefore be used for reference only and the results of dictionary editing can be reflected only in the personal dictionary. That is to say, the contents of the basic dictionary can be modified by the dictionary editor and the results can be stored in the personal dictionary.

When referring to to dictionary entries for translation or dictionary editing purposes. priority is given to the personal dictionary.

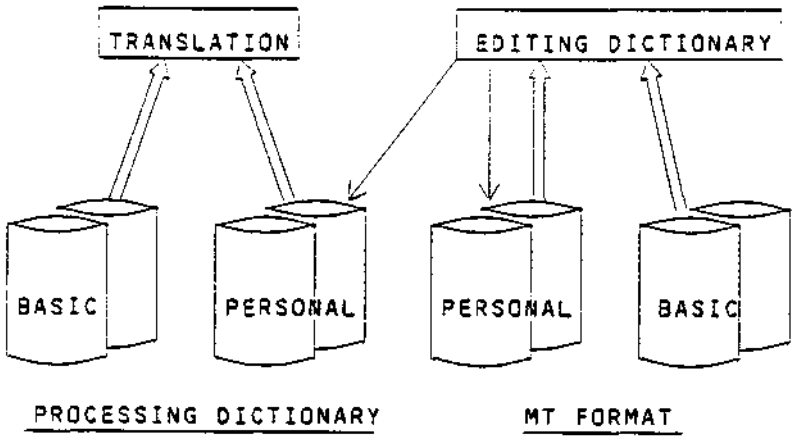

Fig. 4 Flow of Dictionary Data

5.2 Contents of dictionaries

Dictionaries can be divided broadly into two types, that is, unilingual dictionaries such as an English dictionary and Japanese dictionary, and bilingual dictionaries such as a Japanese-English transfer dictionary and English-Japanese transfer dictionary. Unilingual dictionary items include entry information,

morphological information, and syntactic and semantic information. Bilingual dictionary items include entry information, syntactic and semantic information (translation term selection condition), translation entry information, translation syntactic and semantic information. These are further divided based on the parts of speech.

5.3 Dictionary editor

The dictionary editor is a tool to facilitate dictionary editing. But the optimal format and functions of the editor depend on the user's skill in using the tool. To meet different levels of users, the unified system provides two types of editor, allowing users to select one suitable to their levels for dictionary editing purposes.

(1) Table format editor

The MT format for editing is written in the S expression of LISP as shown in Figure 7. This is very difficult to read for general users. Besides, there is a large volume of data to input and the patterns of dictionary contents to write differ with the type of dictionary and with the parts of speech. Moreover, since this translation system requires different dictionaries in different stages translation, it is necessary to give consideration to the influence to all dictionaries when editing a single word The table format editor was developed to solve this problem.

<1> Table format

As shown in Figure 5, this editor expands the contents written in the S expression into a table format. There are as many tables as the number of dictionary content description patterns, which differ with different dictionary types and parts of speech.

It is passible for users to edit a dictionary simply by entering items according to the table displayed without concern for dictionary content patterns which differ with the type of dictionary and the parts of speech.

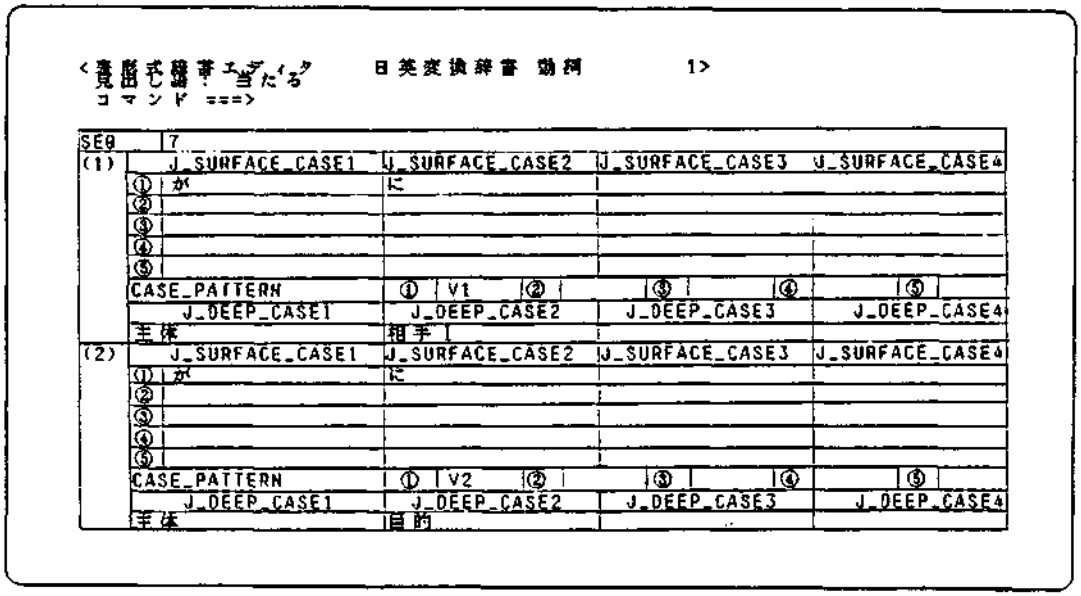

Fig. 5 Table Format on Dictionary Editor (Sample Table for a verb)

<2> Concatenated dictionary editor

Editing one word requires the conformity of three MT formats. If, for example, a modification in the analysis MT format (MT format for analysis dictionary) in the other MT format, a correct translation result cannot be obtained.

The concatenated dictionary editor performs editing according to the procedure shown in Figure 6. The editing operation is transferred from analysis through transfer to generation, thus preventing the omission of editing of single words.

This editor is named the concatenated editor because it performs editing as a consistent flow.

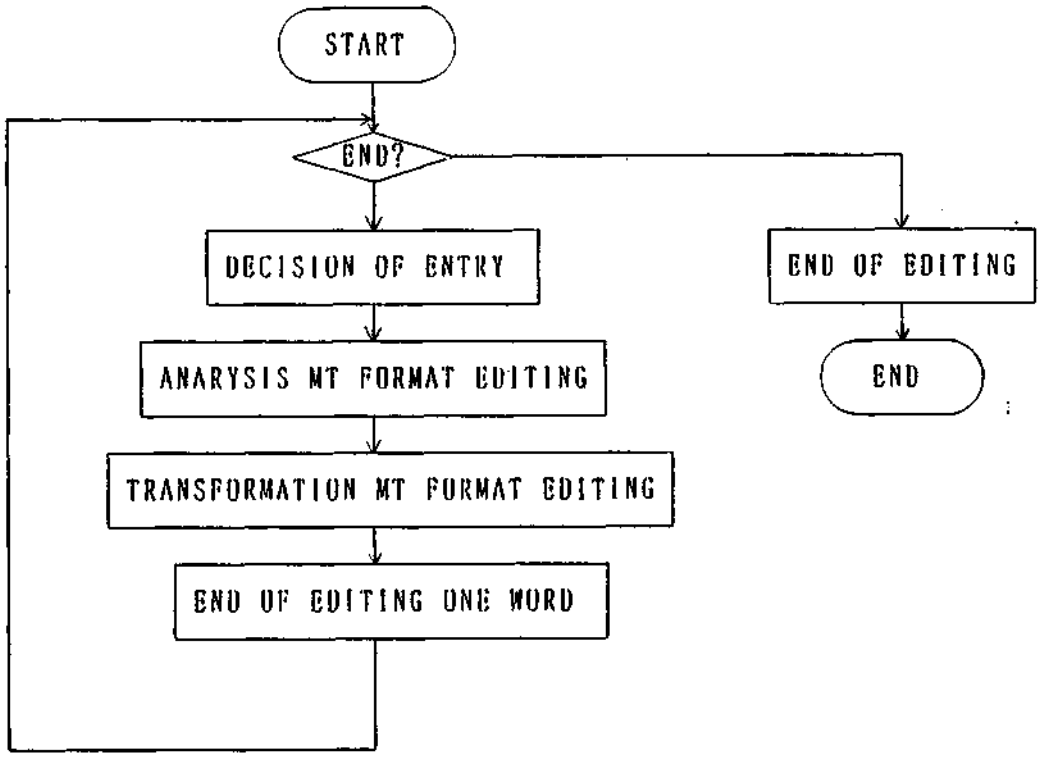

Fig. 6 Editing Procedure of Table Type Dictionary Editor

<3> Retrieval of the dictionary

There are variants, e.g., color and colour. The Japanese language has many variants. It is wasteful to enter all these variants individually in the dictionary. If, therefore, there are several variants, one of them is entered as a dictionary entry and the rest are described as variants in the dictionary.

In dictionary editing, it may be possible that a non-entry word is specified. In such a case, it is inconvenient if the specified word is treated as not entered, even though a variant is entered. To avoid this, when a word is specified, at first, the morphological analysis dictionary is looked up before accessing the MT format. Because it has all variants entries.

<4> Editing function

The following functions are provided to facilitate the use of the editor.

(a) Saving, cancelation, and deletion of results of editing

It is possible to save the result of editing in a personal dictionary, or cancel the editing operation.

(b) Copying

It is possible to copy entries of another dictionary onto the editing screen. This makes it easy to add new entries to the current dictionary. (Since tables differ with different types of dictionary and parts of speech, this function can be used with dictionaries of the same type and words of the same part of speech.)

(c) Reference

It is possible to display entries of another dictionary in the lower part of the display screen by means of a command. This function can be used to look up another dictionary when editing the current dictionary or to refer to another entry of the dictionary.

(d) Return

Editing is performed in the sequence of analysis, transfer, and generation. This function permits returning to the previous dictionary to allow the user to modify it again.

(e) Default

When inputting a new word, the user may be at a loss because there are many dictionary input items, the default values automatically assumed for the essential dictionary input items to allow the user to start his/her translation anyway.

(f) Dictionary management

Sequential numbers are assigned automatically to the entries of the personal dictionary. This helps manage personal dictionary data.

(g) individual editing

It is possible to edit the MT format individually if it is desire to do so.

(2) S-expression dictionary editor

The table format dictionary editor presents the contents of dictionary in an easy-to-see table format. But its flexibility of editing is limited. For users familiar with LISP, it is often more convenient to handle the S-expression than to manipulate the table when making a small modification.

This system therefore provides an S-expression dictionary editor that displays the MT format dictionary directory on the display screen for

editing.

The contents of the dictionary is written in the MT format with the structure of the S-expression of LISP. This editor displays the contents of the dictionary in an easy-to-see form in the data display area of the editing screen (Fig.7). The user can edit the contents of the dictionary in this area. Since there are no restrictions on the length of dictionary items and iterations, editing can be performed with items and iterations, editing can be performed with good flexibility.

With this editor, editing is performed for each dictionary. Users "are responsible for the conformity of dictionaries.

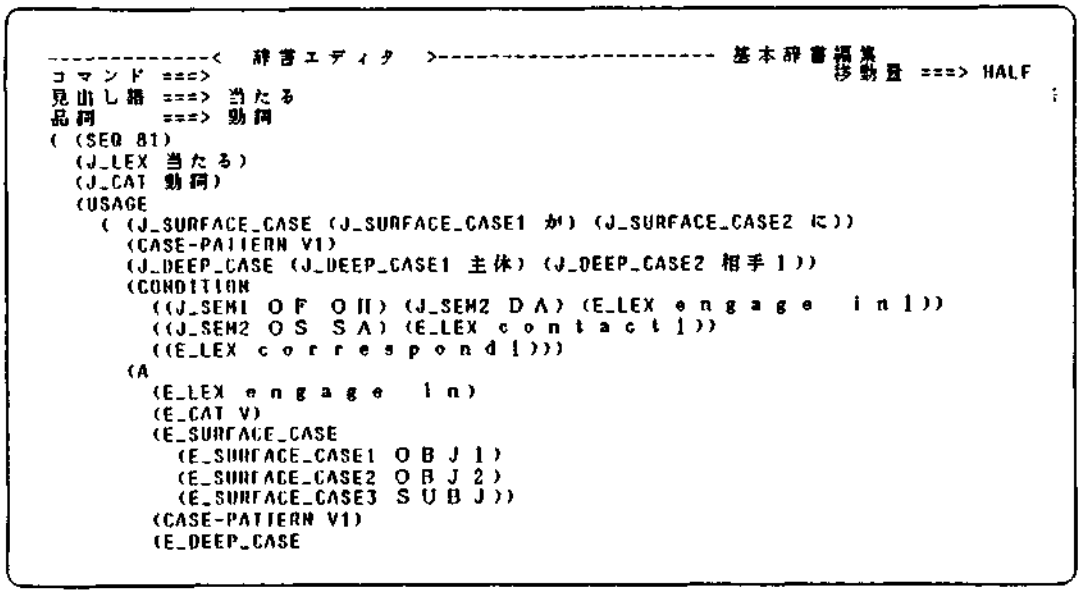

Fig. 7 S Expression on Dictionary Editor Screen

6. Image dictionary

The image dictionary is an electronic dictionary system using a disk which stores the image of the contents of an, ordinary human dictionary. The system is provided with a powerful retrieval function to look up the dictionary and is capable of displaying and printing dictionary information in the form of image.

6.1 Purpose

When using the machine translation system, it is sometimes necessary to look up an ordinary human dictionary:

(1) To perform post-editing

(2) To edit the contents of the dictionary

As a general translation support tool, the machine translation system requires an electronic form of some ordinary dictionaries.

6.2 Structure and function

Figure 8 shows the structure of image dictionary.

The image data (primary information) is the contents of the New Concise English-Japanese Dictionary stored page by page on a disk. The data base is created by the image data accumulation function of the electronic filing system.

The retrieval information (secondary information) is a data base that contains the information necessary to retrieve image data. It is created by the data accumulation function of the interactive information retrieval system. The image data and retrieval information are linked with dictionary pages as keys. If certain secondary information is retrieved by functions of the information retrieval system, the image data linked to it is taken out by functions of the electronic filing system for display and printing. This system has the following functions:

(1) Retrieval

The desired secondary information can be retrieved simply by specifying part of the entry word such as forward correspondence or backward correspondence. It is possible for the user to select the desired element from the retrieved information set and print or display it as images.

If the explanation of a word extends over two or more page, it is possible to display or print the page that contains the desired definition of the entry word or examples of its use by specifying the word and the positioning item.

(2) Display

It is possible to the image data (one page of the dictionary) linked to the retrieved secondary information on the image display. The data can be enlarged or moved upward, downward, leftward and rightward.

(3) Printing

Image data (one page of the dictionary) can be printed on a Japanese line printer.

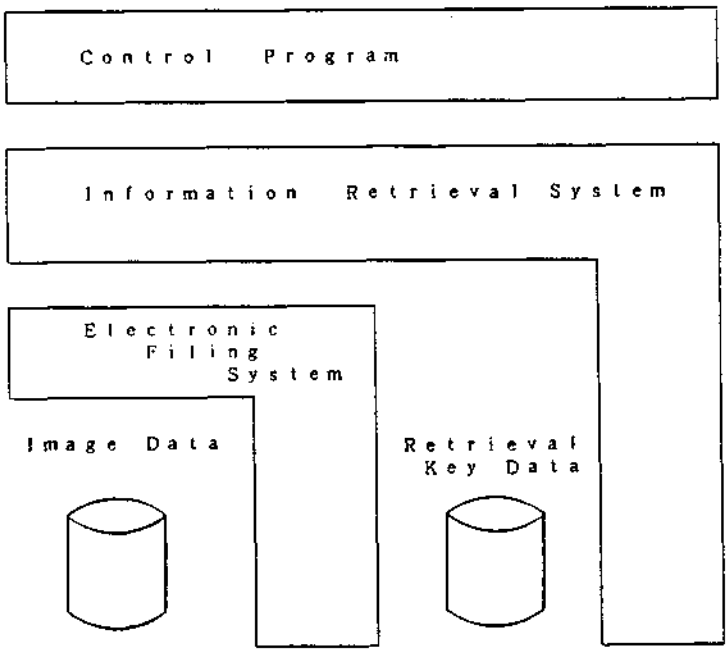

Fig.8 Structure of Image Dictionary

7. Translation experiments and evaluation

This section explains an outline of the translation experiments conducted at RIPS, their results and evaluation.

7.1 Method

To evaluate the system as a machine translation system for practical use, experiments were conducted with the language processing system alone and with the unified system.

7.2 Experiments with the language processing system

Since the language processing system and the unified system were developed in parallel, translation experiments and evaluation of the results were conducted with the language processing system alone. As explained previously, this was an effective method respecting memory capacity and processing speed as well.

7.2.1 Experiment environment

(1) Computer system

* FACOM M380 with MSS (Mass Storage System)
* User's working area : 4MB to 8MB

An experiment can be performed with a minimum area of 4MB. But actual experiments were conducted with a maximum area of 8MB to increase the speed of translation under UTILISP.

(2) Software

OS : OS IV/F4 MSP

Programming language : Kanji version of UTILISP with capabilities of collecting garbage on a disk and accessing VSAM.

(3) File environment

100MB upward for storing the translation programs, input-output texts, grammatical rules, dictionaries and translation utilities on general-purpose files. 200MB upward (VSAM files for the results of each step translation (for examination and dictionaries.

7.2.2 Translation experiments

Japanese into English

3.000 sentences sampled from JICST's literature file. The results of the experiments were evaluated for improvement of the system. Also, 10.000 sentences were translated once. The results of the experiment were evaluated for improvement of the system.

English into Japanese

3.000 sentences sampled from INSPEC's files. The results of the experiments were evaluated for improvement of the system.

7.2.3 Results of morphological analysis

Japanese into English

Japanese is an agglutinative language. This makes it difficult to identify each word in one sentence. In this system, a sentence is analyzed by looking up the dictionary for the character string. The experiment found that the error rate

in identifying words was relatively high for kanji and words of foreign origin expressed in katakana. This is because these words are not entered in the dictionary of nouns. What affect the subsequent process of translation, e.g., syntax analysis are unentered declinable words and wrong recognition of case of particles. These were 542 errors in the translation of the 10,000 sentences.

English into Japanese

The English morphological analysis also identified words based on the dictionary as in the case of Japanese. The English analysis was much more accurate than the Japanese morphological analysis. Errors were proper nouns, hyphenated compound nouns and abbreviations. Such errors seldom affect the syntax analysis.

7.2.4 Result of translation

Japanese into English

The translation was carried out in four steps, that is, morphological analysis, syntax analysis, syntax transfer, and generation. It took a turnaround time of three months to translate 10,000 sentences on a batch basis. This was because it took about ten seconds to translate one sentence. One second of CPU time needs an elapse time of 60 seconds. The rate of the successful termination of processing was measured for each of the three steps; analysis, transfer, and generation. The average successful rate was from 98% to 94%. The translated sentences are now under evaluation by experts.

English into Japanese

This is now under way. The results, together with the evaluation of the Japanese-English translation, will be announced at the meeting.

7.3 Unified system experiment

7.3.1 Experiment environment

Translation experiments were carried out with the unified system incorporating the language processing system. The experiment environment was the same as that for the language processing system with the following exceptional;

(1) User's working area

* Unified system : 3.6 to 4MB (Within 4MB for language processing system)

7.3.2 Translation experiment

A translation experiment was conducted for 45 sentences sampled from sentences whose translation by the language processing system was carried out successfully.

(1) Translation speed

Figure 9 shows the correlation between the length of a sentence and the speed of translation.

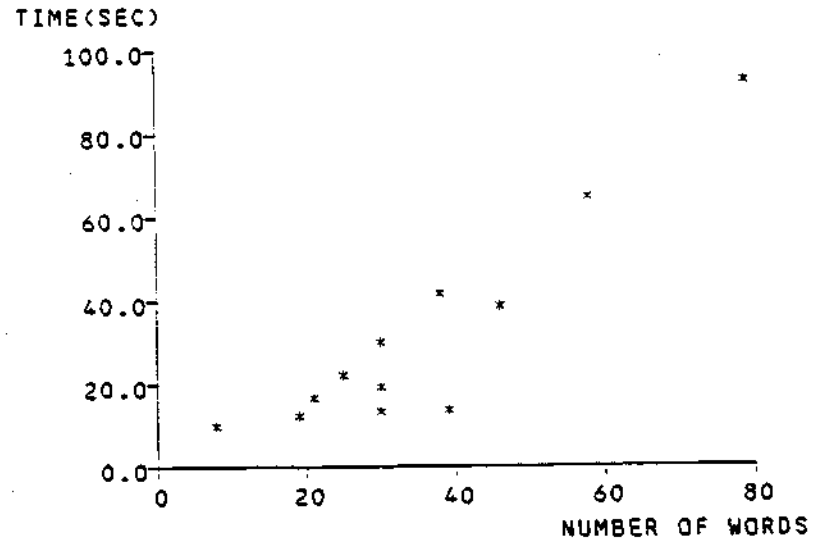

Fig.9 Correlation between Number of Words and the Speed of Translation

Generally, the longer a sentence, the longer the time of translation. But there were cases in which the time required for translation varied with sentences of the same length as shown in Table 1. The time for generation was nearly the same, but there were differences in the times required for analysis and transfer. Differences in the time for analysis, which accounted for from 65% to 80% of the total translation time, were particularly significant. This shows that the time required for translation affected by the complexity of a sentence as well as its length. For example, (i) no nominative, (ii) the word having multiple meanings, (iii) metaphorical expression.

Table 1 Speed to Translation Sentences of the Same Length

NUMBERS OF WORD	ANALYSIS STEP SEC	TRANSFER STEP SEC	GENERATION STEP SEC	TOTAL SEC
30	8.73	0.76	3.84	13.33
30	10.67	1.11	3.80	15.58
30	24.38	2.25	3.14	29.77

(2) Text editor

The translation experiment with the unified system has identified the following problems with respect to man-machine interface:

* It takes much time to initialize the translation editor.
* Kana-kanji conversion for Japanese sentence input involves frequent key operations.

8. Conclusion

The unified system has satisfactory performance as a translation system for practical use. But there still remain a lot of problems to be solved, such as unsatisfactory experiment environments, inadequate scale of translation experiment, and insufficient time for evaluation and analysis of the results of the experiments.

The following are some of the problem areas identified in the experiments:

(1) The language processing system employs LISP for system flexibility and ease of development. The problems involved in the use of this language include the standardization of grammatical description and exclusion of near non-grammatical texts.

(2) The unified system uses an 8MB area on a large computer (equivalent to 16MIPS machine) for translation processing. But an area about three times the present size is required for practical use. In addition, the disk capacity and access speed are required to be improved.

(3) For users convenience, it is necessary to provide the terminal with multiwindow capability and develop powerful text and dictionary editors.

Acknowledgments: The authors were given help and advice from Prof. Nagao (Kyoto Univ.) about promotion of Mu-project, and from Prof. Tsujii and Dr. Nakamura (Kyoto Univ.) about interface between language processing system and the unified system. We appreciate them. The authors thank those who gave us valuable guidance and advice and took the trouble to prepare the software and dictionaries. Our gratitude also goes to the Science and Technology Agency and the Ministry of International Trade and Industry for their assistance to our work.

Reference:

In English

(1) Nagao M.: Structural Translation in the Generation Stage of Mu Japanese to English Machine Translation System, Theoretical and Methodological Issues in Machine Translation of Natural Language, Hamilton, New York, 1985.

(2) Sakamoto Y., Satoh M. and Ishikawa T.: Lexicon Features for Japanese Syntactic Analysis in Mu-project-JE, Coling4, Stanford, 1984.

In Japanese

(3) Sakamoto Y. and Ariga T.: Basic Design for Unified System in Mu-project. Working Group Report of Natural Language Processing (WGNL) 46-6, Information Processing Society of Japan, 1984.

(4) Nagao M., Tsujii J., Nakamura J., Sakamoto Y., Toriumi T. and Satoh M.: Outline of Machine Translation Project of Science and Technology Agency, Journal of Information Processing Society of Japan, vol. 26, no. 10, 1985.