AVAILABILITY OF ENGLISH-JAPANESE MACHINE TRANSLATION USING A PERSONAL COMPUTER

Hiroyasu NAKASA and Yoshiyuki KANAMARU
Komae Research Laboratory
Central Research Institute of Electric Power Industry
Komae-shi, Tokyo, JAPAN

ABSTRACT

Several software programs available for English-Japanese machine translation have been evaluated in various types of basic and practical sentences. The results are unfortunately dissatisfactory because of mainly the weakness in grammatical bases. However, one of the programs is open to the public in the BASIC language for a personal computer, and it turns out that powerful translation is possible after improvements to solve grammatical problems and to grade up the dictionary.

This paper presents the processing method for English-Japanese translation of the modified program and demonstrates the resultant high potentiality of the translation ability.

INTRODUCTION

The needs for English-Japanese machine translation are growing up in order to reduce the labor time and cost for the translation of a large number of English technical papers. On the other hand, great progress has been made in the hardware of a personal computer, especially in the capacity of computer memory and the speed of data processing. Then, it is to be desired that such a remarkable progress should be reflected in the application field of machine translation.

This paper presents the result of evaluating English-Japanese translation programs available and discusses the practical performance of the program for a personal computer.

TRANSLATION PERFORMANCE EVALUATION TESTS

Recently, several software programs become available to English-Japanese machine translation for personal computers [1, 2] as well as large computer systems. Their practical performance for translation of technical papers has been evaluated using various types of test sentences. The result was unfortunately dissatisfactory mainly because of the weakness in grammatical bases of those programs.

Table 1 shows typical examples of the results in the evaluation test using various types of sentences. The lengths of the sentences shown in the lower part are only about 10 words. However, even the translation program for a large computer does not have a perfect ability to translate such English sentences, and also other programs for the personal computer do not get good remarks.

One of the translation programs evaluated is written in the BASIC language for a personal computer, and it is open to the public [1]. This program has merely a poor and crude ability of machine translation as it is, but it has the ability to unlimitedly raise the practical performance.

This English-Japanese translation system consists of a main program with only less than 1000 lines and four kinds of data-file modules: "E-J dictionary", "appropriate-word selection", "grammatical rules for sentence construction" and "Japanese polishing". These system components are organically connected with each others, but they are able to independently modify or add supplementary rules.

There was special attraction in the task of the program improvement using an 8-bit personal computer (PC-8001mkII/NEC ‹clock: 4 MHz, memory size: 64 KB›) with 2D floppy disks. Resultingly, the usefulness of powerful translation has been demonstrated after comprehensive and cautious improvements to solve grammatical problems and to give a good collection of expression to the dictionary.

Table 1. Typical Examples of Evaluation Test Results on the Machine Translation Ability.

Sentence No	A Large-computer	B PC/Machine	C PC/BASIC	D PC/BASIC
1	○	△	○	×
2	○	×	×	×
3	×	×	×	△
4	△	△	△	○
5	△	△	×	○
6	○	×	×	×
7	○	×	×	×
8	△	×	×	×
9	○	×	△	△
10	△	×	×	×

○ : good
△ : the dictionary should be changed.
× : grammatical revision is needed.

Examples of test sentences

(1) Any dirt or other impurities will decrease the engine efficiency.
(2) It has a one-year guarantee against mechanical defects.
(3) Chemical reactions are the processes that convert substances into other substances.
(4) Acid rain is one of the big pollution problems.
(5) Find the common factor of 32 and 8.
(6) Explain why water is effective in stopping some fires.
(7) A major feature is the ability to identify 'patterns' of information.
(8) The electric charge of nuclei is positive, and that of electrons is negative.
(9) Relays minimize service interruptions and damage due to abnormal conditions in the system.
(10) This door is designed to close to obtain comfortable room temperature.

TRANSLATION PROCESSING OF THE MODIFIED PROGRAM

The translation processing of this modified program is similar to the original one, as follows:

(1) input of an original sentence written in English.
(2) sentence disintegration into words or idiomatic phrases.
(3) consulting the "E-J dictionary" file and the "appropriate-word selection" file for a word with many meanings.
(4) selection of Japanese words with proper meanings, and determination of a numerical code for a part of speech.
(5) generation of a numerical code progression which presents the order of Japanese words composition.
(6) reduction of the length of the code progression using the "grammatical rules for sentence construction" file.
(7) polishing up the sentence using the "Japanese polishing" file, and construction of the Japanese sentence.

Figure 1 and 2 show typical examples of the translation processing of this modified program.

Figure 3 shows an example of consulting the "E-J dictionary" file for a phrase of "convert # into >", where 4 kinds of symbols can be used as a substitute word, and also 2 kinds of codes are used for both the part of speech and the word attribution.

The "appropriate-word selection" file for the word with many meanings performs the selection of the Japanese word with a proper meaning and the determination of the numerical code related to the part of speech. Figure 4 shows the case of the word, "that", which has two kinds of meanings of the pronoun and the article, and the word has the code of "r" (relative pronoun). According to the rule that if the left-side word is 1 (noun) or 3 (verb), and if the word concerned includes the code of "r", then 8 (article) is deleted and 2 (pronoun) is adopted, the meaning of "that" is fixed as the pronoun, and the numerical code is fixed as "2".

After the code number of each word is fixed, a numerical code progression is generated to present the order of Japanese words composition. It becomes possible with the "grammatical rules for sentence construction" file to make reduction of the progression length and construction of the Japanese sentence.

Hydrogen combines with oxygen to form water.

"E-J dictionary"

```
hydrogen        1水素
combines with   3と結合する
oxygen          1酸素
to              7のために  T
                7へ     g
form            3を形成する
water           1水     u
                0
```

"appropriate-word selection"

```
7のために   7へ
        77:     77;t;R<>tn;RR<>tn;
                77;h;R<>h;RR<>h;
                77;g;R<>g;
```

"grammatical rules for sentence construction"

```
* 1317310
  3と結合する
  1酸素
  31;3;1;T1<>&;T2<>r;T0<>j;J0<>>;J2<>3;J2<>4;J0#VRB;J1#OBJ;
  3酸素と結合する
* 137310
  3を形成する
  1水
  31;3;1;T1<>&;T2<>r;T0<>j;J0<>>;J2<>3;J2<>4;J0#VRB;J1#OBJ;
  3水を形成する
* 13730
  3酸素と結合する
  7のために
  3水を形成する
  373;3;0;T-1<>1;T-1<>v;T0<>T;T1=T;J1#N;J0#+と;
  3酸素と結合すると水を形成する
* 130
  1水素
  3酸素と結合すると水を形成する
  13;X;0;J-1<>>;T1<>v;J1<>>;J0#SBJ;J0#+は;
  X水素は酸素と結合すると水を形成する
* X0
```

"Japanese polishing"

水素は酸素と結合すると水を形成する。

Fig.1 Typical Example of Translation Processing (1).

Figure 5 shows an example of the process to change "373" to "3" and it shows various symbols and operation parameters.

This "grammatical rules for sentence construction" file is the core of the translation processing, and it consists of the following major flow of grammatical operation: noun (article, adjective, conjunction) -> preposition -> adverb -> verb -> auxiliary verb -> clause -> sentence.

Chemical reactions are the processes that convert substances into other substances.

"E-J dictionary"

```
chemical            5化学的
reactions           1反応
are                 3である v
the                 8
processes           1プロセス
that                2それ    or
                    8あの
convert # into >    3#を>に変換する
other               1他のもの
                    5その他の
substances          1物質
.                   0
#1                  0       #1
                    1物質
```

"appropriate-word selection"

```
2それ    8あの
          28:       28;8;L=13;M=r;
1他のもの   5その他の
          15:       15;1;R=158;M<>a;
```

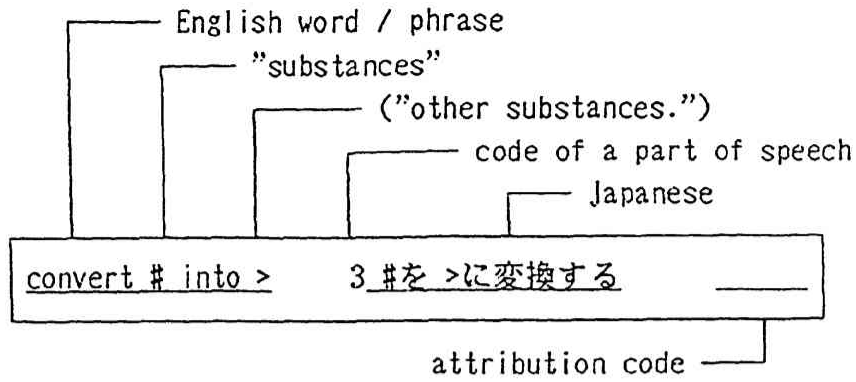
"grammatical rules for sentence construction"

```
* 513812351001
  8
  1プロセス
  81;1;0;
  1プロセス
* 51312351001
  5化学的
  1反応
  51;1;0;J0<>;
  1化学的反応
* 1312351001
  5その他の
  1物質
  51;1;0;J0<>;
  1その他の物質
* 131231001
  1化学的反応
  3である
  1プロセス
  2それ[プロセス、反応]
  3#1を>に変換する
  13123;X;0;T3=r;J0#SBJ;J1#VRB;J2#OBJ;J3#N;J0#+は;J1<->J4;
  X化学的反応は#1を>に変換するプロセスである
* X1001
```

"Japanese polishing"

化学的反応は物質をその他の物質に変換するプロセスである。

Fig.2 Typical Example of Translation Processing (2).



Attribution code

a :	at the beginning of the sentence
f :	future
g :	place
h :	human
i :	incomplete transitive verbs
j :	intransitive verbs
k :	perception verbs
m :	physiological words
n :	numerical words
o :	objective of pronouns
p :	past tense
q :	interrogatives
r :	relatives
s :	mental words
t :	time
v :	"be" verbs
w :	weather
C :	comparative degree
E :	"there be"
F :	"and", "or", "but"
G :	gerund
H :	"have", "let"
I :	"it"
K :	subjunctive words
M :	"must"
N :	negative words
P :	past participle
T :	infinitive
U :	superlative degree

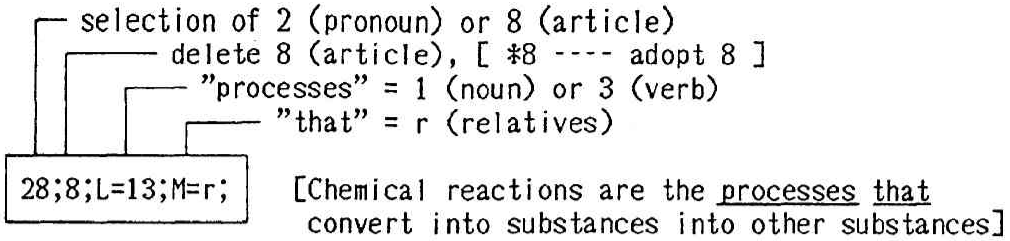
Symbols used in the "E-J dictionary" file

# :	substitute words
> :	right-side words
< :	a left-side word (for Japanese only)
$:	a right-side word (for Japanese only)

Numerical code related to a part of speech

0 :	"." or ","
1 :	noun
2 :	pronoun
3 :	verb
4 :	auxiliary verb
5 :	adjective
6 :	adverb
7 :	preposition
8 :	article
9 :	conjunction
X :	interjection or clause

Fig.3 Processing in the English-Japanese Dictionary.



Symbols used in the "appropriate-word selection" file

* :	adoption (↔'no mark': deletion)
= :	necessary condition
<> :	prohibitive condition
FT :	at the biginning of the sentence
LT :	at the end of the sentence
LL :	the double left-side word
L :	the left-side word
M :	the object word
R :	the right-side word
RR :	the double right-side words

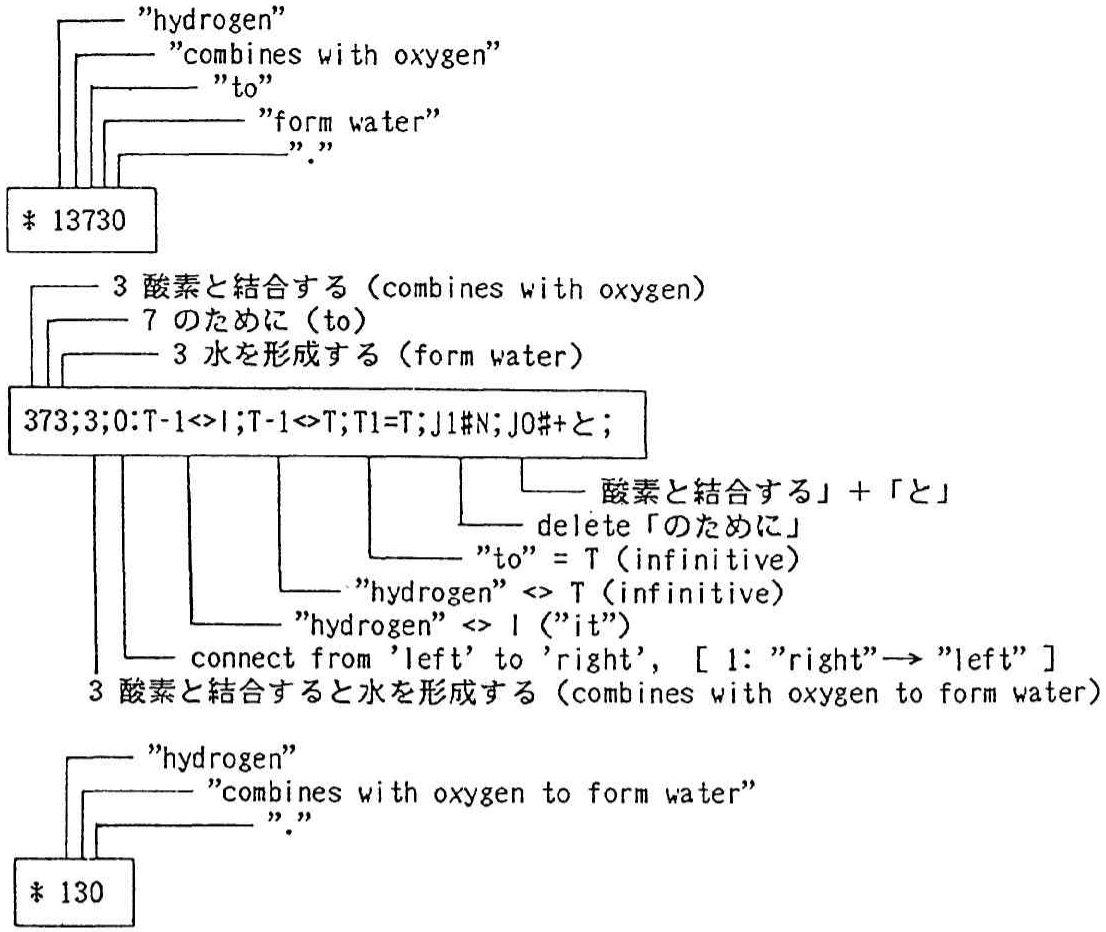
Fig.4 Processing in the Appropriate-Word Selection.

Through these processing, it is possible to compile a combined-words sentence and to polish up it using the "Japanese polishing" file, and to construct the Japanese sentence.

AVAILABILITY OF THE MODIFIED PROGRAM

At present, the achievement level of the English-Japanese translation program is not so high yet, but the prospects for the enhancement of translation performance is good through grading up both the grammatical basis and the dictionary.

The problem of the translation speed can be easily solved by means of the hardware change from an 8-bit personal computer to a 16- or 32-bit one with a hard disk. Table 2 shows the result of measuring the processing time to translate the same sentences as shown in Table 1, using 4 kinds of systems with an 8- or 16-bit personal computer plus floppy or hard disks and a program with or without the "KANJI" BASIC or the BASIC compiler. As seen in Table 2, the problem of the translation speed can be solved using a 16-bit computer with a hard desk and using the BASIC compiler.



Symbols in the "grammatical rules for sentence construction" file

J :	Japanese
T :	attribution code
= :	necessary condition
<> :	prohibitive condition
# :	operation to Japanese
N :	deletion of Japanese
+ :	addition
- :	deletion
<-> :	order change of Japanese words
>>> :	remove of a particle (「を」, 「に」, 「と」, 「が」)
#SBJ :	memorizing as Subject-words
#OBJ :	memorizing as Object-words
#VRB :	memorizing as Predicate-verb
KAKO :	Past Tense
ONBIN :	Euphonic Conjunction
UKEMI :	Passive Voice
CONJ :	Conjunction Phrase
HITEI :	Negative Sentence
MUST :	"must"
SIEKI :	Causation
ROOT :	Root Form
MEIR :	Imperative Mood

Fig.5 Processing in the Gramatical Rules for Sentence Construction.

Table 2. Processing Time (sec) to Translate an Input Sentence.

Sentence No	8-bit PC K-BASIC	8-bit PC BASIC	16-bit PC K-BASIC	16-bit PC+HD K-BASIC-C
1	287	153	11	3
2	254	140	14	3
3	314	173	13	3
4	257	157	10	3
5	225	125	8	2
6	316	179	14	4
7	290	178	12	4
8	340	203	15	5
9	362	198	15	5
10	267	153	11	3

On the other hand, the fully automatic operation without any personnel is not practical even in the future of machine translation, because of various problems such as the presence of input errors and the shortage of registered words. These facts may indicate the advantage of the use of the personal computer instead of the large computer system.

CONCLUSION

In conclusion, the availability of the English-Japanese machine translation using a personal computer is sufficiently high in a limited application field of technology due to the excellent expandability of the program described here. It is possible in future to strengthen the machine translation ability through piling up experiences to translate as many typical sentences as possible.

REFERENCES

1) T. Ueno: "Machine Translation Programming Work Using a Personal Computer" (in Japanese)", [RASSERU Co.], (1986).
2) K. Shibata: "PC-9800 (MS-DOS) Machine Translation Programs" (in Japanese), PC Magazine 1986/12 〈No. 53〉, 68-83.